

# INTERDISCIPLINARY APPROACH TO THE LUBRICATION OF CONCENTRATED CONTACTS

## CASE FILE COPY

A symposium held at  
RENSSELAER POLYTECHNIC INSTITUTE  
Troy, New York  
July 15-17, 1969



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION



# INTERDISCIPLINARY APPROACH TO THE LUBRICATION OF CONCENTRATED CONTACTS

Proceedings of a NASA-sponsored symposium held  
July 15-17, 1969, in Troy, New York

Edited by P. M. Ku  
*Southwest Research Institute*



*Scientific and Technical Information Division*  
OFFICE OF TECHNOLOGY UTILIZATION  
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION  
1970  
*Washington, D.C.*

---

For Sale by the Superintendent of Documents,  
U.S. Government Printing Office, Washington, D.C. 20402

Price \$2.50 (paper cover)

*Library of Congress Catalog Card Number 70-606643*



## Foreword

THESE PROCEEDINGS ARE the record of some very lively discussions on the topic "Interdisciplinary Approach to the Lubrication of Concentrated Contacts—A NASA Symposium" held July 15–17, 1969 at Rensselaer Polytechnic Institute, Troy, New York. It is hoped the printed pages will be a considerable contribution in this field, even though the printed word cannot truly reflect the interest and enthusiasm shown at the symposium. The NASA objectives, in presenting this symposium as a part of a series of interdisciplinary symposia, are to enkindle interest of those presently outside the field in the various problems within this field. Such problems include, among others for example, lubrication, wear, fretting, surface damage and fatigue. A complete understanding of the basic mechanisms involved in such phenomena can always profit from different points of view on the various phenomena. It is here that the interdisciplinary approach may be so useful toward finding ultimate solutions. Different points of view are not only welcome, they are solicited. Most certainly, a multidisciplinary approach cannot help but illuminate some of the mysteries with which we are plagued at present.

EDMOND E. BISSON, *Associate Chief,*  
Fluid System Components Division  
NASA Lewis Research Center

## Steering Committee

P. M. Ku, Southwest Research Institute (Chairman)

G. S. Ansell, Rensselaer Polytechnic Institute

E. E. Bisson, National Aeronautics and Space Administration

K. E. Demorest, National Aeronautics and Space Administration

G. C. Deutsch, National Aeronautics and Space Administration

D. G. Flom, General Electric Company

E. E. Klaus, Pennsylvania State University

M. C. Shaw, Carnegie-Mellon University

R. P. Shevchenko, Pratt and Whitney Aircraft



## Preface

IN AN EFFORT to further the advances in the complex and multidisciplinary subject of lubrication, the National Aeronautics and Space Administration (NASA) inaugurated in 1967 an interdisciplinary lubrication symposium series. The first meeting of the series, held in San Antonio, Texas, on November 28-30, 1967, was devoted to a critical appraisal of the level of understanding and the needed research in the area of sliding friction and wear under unlubricated and boundary-lubrication conditions.\* The second meeting, held in Cleveland, Ohio, on November 19-21, 1968, again emphasized unlubricated and boundary-lubricated sliding friction and wear but attempted to focus attention on some selected theoretical and practical problems by discussing in small working groups their implications and possible methods of attack.\*\*

Having thus provided for an examination of sliding friction and wear under unlubricated and boundary-lubrication conditions in considerable breadth and depth, the Steering Committee recommended to NASA in September 1968 that further attention be turned to an evaluation of the collective understanding and the needed research in the lubrication of highly loaded machine elements in a combined rolling and sliding situation. The recommendation was accepted by NASA and, accordingly, a symposium entitled *Interdisciplinary Approach to the Lubrication of Concentrated Contacts* was held on the grounds of Rensselaer Polytechnic Institute, Troy, New York, on July 15-17, 1969.

As in the case of the two previous meetings, this symposium again employed the technique of dividing the whole area into a number of relatively manageable topics, with each topic introduced by an invited lecture followed by general discussions. Moreover, interdisciplinary approach was emphasized in the selection of the invited lecturers and of the rest of the invited participants. Of the 94 persons who accepted the invitation to participate in this symposium, about 60 percent were lubrication research engineers, 15 percent were basic scientists, and 25 percent were design and development engineers.

---

\* Ku, P. M., ed.: *Interdisciplinary Approach to Friction and Wear*. NASA SP-181, 1968.

\*\* Bisson, E. E.; and Ku, P. M., eds.: *Friction and Wear Interdisciplinary Workshop*. NASA TM X-52748, 1970.

To ensure adequate and meaningful discussions, the lectures were preprinted and mailed to each preregistered participant 1 month prior to the meeting. During the meeting each topic was allotted an equal division of time between the lecture and the discussions.

The invited lectures were all over 6 months in preparation. After review by the Steering Committee and appropriate revisions by the lecturers, 14 of the 16 lectures were issued as preprints. The two lectures not available for preprinting were "The Structure of Solids" by H. G. F. Wilsdorf (University of Virginia, Charlottesville, Virginia) and "Effect of Materials—Materials Science Viewpoint" by S. Eisner (Norton Company, Troy, New York). Professor Wilsdorf and Mr. Eisner presented their respective lectures orally at the meeting; both of which were well received. Regrettably their lectures are still not available for publication.

For various reasons, three of the lecturers were unable to be present at the meeting to deliver their lectures in person. They were D. Dowson (The University of Leeds, Leeds, England), whose lecture "Elastohydrodynamic Lubrication" was given by A. Dyson (Shell Research Limited, Thornton Research Center, Chester, England); H. Blok, whose lecture "The Postulate about the Constancy of Scoring Temperature" was given by P. M. Ku (Southwest Research Institute, San Antonio, Texas); and H. Eyring (University of Utah, Salt Lake City, Utah), whose lecture "The Structure of Liquids" was given by K. Liang (Eastman Kodak Company, Rochester, New York). However, as in the other cases, written discussions on these lectures were submitted to the respective lecturers after the meeting, and the closures were written by the lecturers.

It should be mentioned here that the discussions that actually took place during the three-day symposium were very lively and provocative. All participants were requested to submit in writing their discussions or comments on any of the lectures for publication and for the lecturers' rebuttal. Unfortunately, in spite of the lively discussions at the meeting, not many written discussions were received. Accordingly, the discussions published herein represent only a fraction of those worthy of record.

I should like to acknowledge here my personal indebtedness to the members of the Steering Committee for their guidance in the planning of the program, for their review of the invited lectures, and for their assistance that contributed toward the success of the symposium.

On behalf of the Steering Committee, I should like to thank the Rensselaer Polytechnic Institute for hosting this symposium. Above all, to all lecturers and discussors without whose active support the symposium could not possibly have come into being, our sincere thanks are due.

P. M. KU  
San Antonio, Texas  
March 1, 1970



# Contents

LUBRICATION OF CONCENTRATED CONTACTS—THE PRACTICAL PROBLEM by B. W. Kelley.....	1
ELASTOHYDRODYNAMIC LUBRICATION by D. Dowson.....	27
MICROSLIP AND CREEP IN CONTACTS by H. Poritsky.....	77
THE POSTULATE ABOUT THE CONSTANCY OF SCORING TEMPERATURE by Harmen Blok.....	153
THE STRUCTURE OF LIQUIDS by H. Eyring and M. S. Jhon.....	249
THE RHEOLOGICAL BEHAVIOR OF LUBRICANTS by H. Naylor.....	279
THE MECHANISM OF CONTACT FATIGUE by W. E. Littmann.....	309
EFFECT OF MATERIALS—GENERAL BACKGROUND by E. V. Zaretsky and W. J. Anderson..	379
EFFECT OF MATERIALS—METALLURGY VIEWPOINT by E. N. Bamberger.....	409
EFFECT OF LUBRICANTS—GENERAL BACKGROUND by J. K. Appeldoorn.....	439
ADSORPTION AND SURFACE ENERGETICS by A. J. Haltner.....	463
CHEMISTRY IN CONCENTRATED-CONJUNCTION LUBRICATION by R. S. Fein.....	489
BEARING DESIGN CONSIDERATIONS by F. W. Wellons and T. A. Harris.....	529
GEAR DESIGN CONSIDERATIONS by W. Coleman.....	551

# Lubrication of Concentrated Contacts— The Practical Problem

**B. W. KELLEY**  
**Caterpillar Tractor Company**  
**Peoria, Illinois**

The manufacturer of machines that depend on the reliable performance of heavily loaded concentrated contacts such as those occurring on gears, bearings, and cams is faced with the development of practical design criteria for predicting the onset of the many potential modes of failure. Almost all of these modes are significantly affected by the material and the lubricant rheology and chemistry. This paper summarizes the plight of the manufacturer and designer in extracting useful information from this multidisciplinary field and attempting to focus attention on those complex problems that have not yet yielded, in any quantitative sense, to research.

**R**ECOGNITION OF CONCENTRATED-CONTACT PROBLEMS probably occurred not long after the first application of gears by Aristotle about 384 B.C. (ref. 1). Rolling-element bearings appeared much later and, while seeming to be a direct plagiarism from Leonardo da Vinci a full two centuries before (ref. 2), a gentleman by the name of Jacob Rowe, Esq., applied for a patent in 1734 (fig. 1) wherein he defined the advantages of such bearings (ref. 3). Rowe stated that with the adoption of such bearings to . . . "wheel carriages . . . one Horse now will do the labour of two . . . and . . . I will suppose that there will be occasion to employ only twenty thousand Horses . . . instead of the 40,000 existing in the United Kingdom, at an annual savings of 1,095,000 pounds per year." Insofar as the writer has been unable to find historical evidence of such sudden prosperity following that statement, it must be assumed that either problems associated with concentrated contact were uncovered, or it was found that just as much cost was absorbed by a horse that was not working as for one that was.

Since that time the rolling element bearing, the gear, and the cam and tappet have yielded to experimental development to a state highly satisfying to the customer. However, as a result of future requirements, today's research engineer or scientist employed by an aggressive industrial manufacturer is faced with two basic problems that frequently drive his



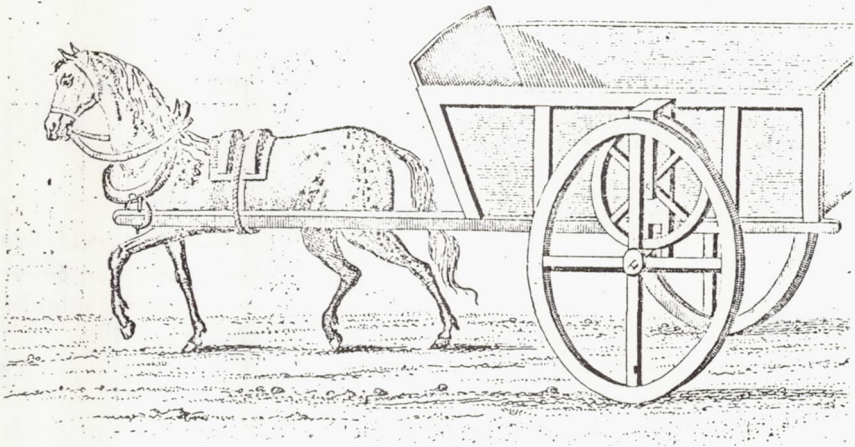


FIGURE 1.—First use of rolling-element contact to reduce axle friction.

more academic colleagues back to the universities. The first is the continual need to develop or conceive well-defined and practical design criteria for such machine elements, and the second is to span the increasing chasm between the findings of basic research and application.

One finds in the fields of structures, mechanisms, component dynamics, and others a reasonably straight-forward approach for the direction that engineering analysis should take and from which research can be conducted in an orderly or logical manner. This is rarely so with the lubrication of concentrated contacts.

While one cannot argue with the state of mathematical elegance achieved in the principles of thick-film lubrication, these solutions generally do not fulfill the requirements of the designer by providing him with an engineering definition of failure. He must fall back frequently on the tribological judgment of "experts." The reason, of course, lies in the staggering multidisciplinary aspects of such problems. One cannot overlook the chemistry, metallurgy, thermodynamics, stress analysis, or physics of the problem of lubricated machine elements and expect to reach a logical solution to even the simplest problem. This same complexity leads to great difficulty in the design of meaningful dynamic experiments, and the lack of a multidisciplinary approach frequently causes the researcher to misinterpret or overlook implications present in his own data.

The importance of the lubricated concentrated contact is brought into focus by exploring the elements of a modern power transmitting machine. Figure 2 shows a powershift transmission for an earthmoving vehicle;

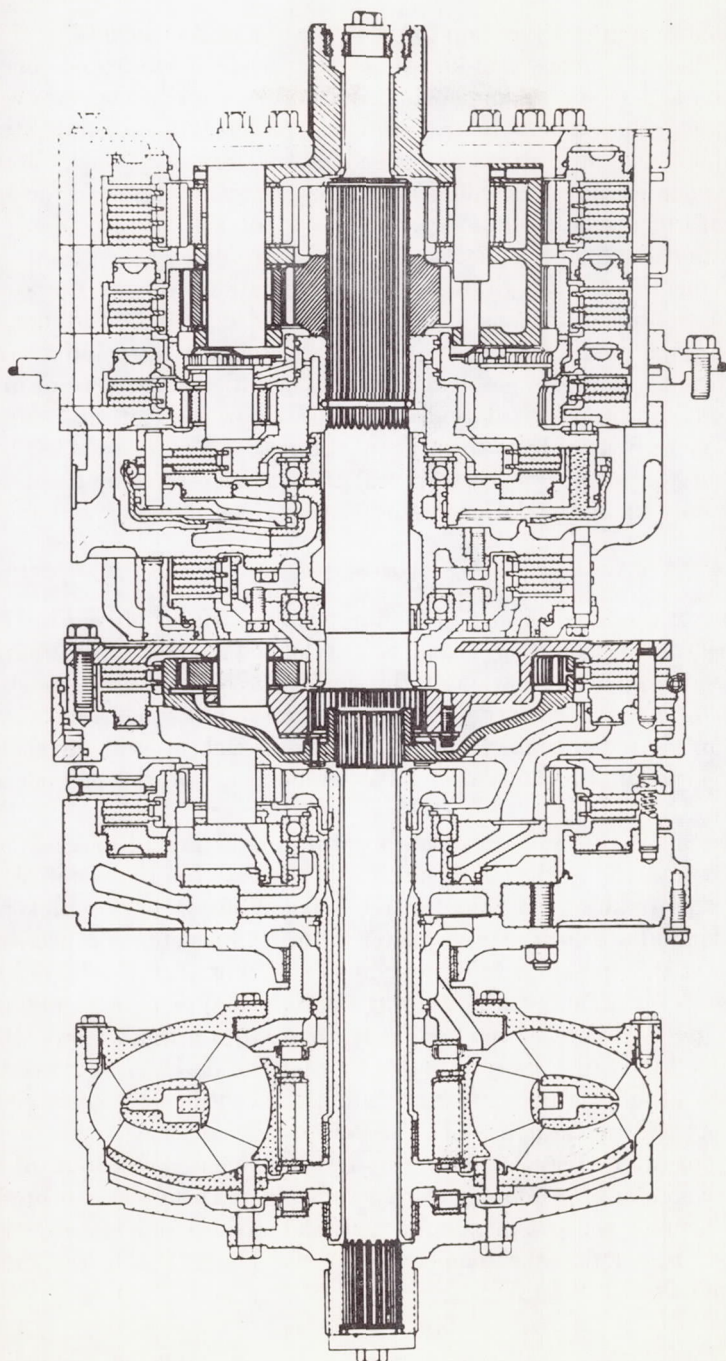


FIGURE 2.—Power shift transmission for earthmoving vehicle.



figure 3 shows a final drive train to the rotors of a large helicopter (ref. 4). Both of these machines contain literally hundreds of lubricated concentrated contacts in sliding and rolling. Both are expected to operate without failure for thousands of hours at full horsepower. Bulk oil temperatures and component temperatures are considerably in excess of what the suppliers of such components would like to see. Both are designed for optimum performance in order to save weight and/or cost.

From the standpoint of mechanics, the design of the gears, cams, and bearings has reached a high degree of sophistication over the years with the computer-aided analysis of deflections, load distribution, dynamics, and stress analysis. The largest problem remaining is in gaining a real insight into the complex problem of their lubrication to prevent or minimize the following principal modes of failure: scoring, wear, and contact fatigue. You will learn from the other chapters that real progress has been made, but the author hopes to reveal the distance remaining between what is now known and the requirements of the designer and manufacturer.

#### SCORING

Scoring or scuffing is a mode of failure on the surfaces of gear teeth or high-performance cams (fig. 4) and is best defined as the transient seizure causing subsequent mutilation of the surface. A classic macroscopic appearance of a scored surface is shown in figure 5, but other appearances may occur dependent upon the material, surface velocities at the time of scoring, type of lubricant, and bulk temperature of the components (ref. 5).

Almost all engineers responsible for the proper design of gears and cams now use the critical temperature hypothesis by Blok (ref. 6) to predict such failure with "nonreactive" oils. While Professor Blok provides a sound basis for the theory later in this symposium, it is necessary that the author preempt his discussion by saying that the hypothesis simply states that failure will occur at a certain conjunction temperature for any given "nonreactive" lubricant. While the critical temperature hypothesis has been a most valuable tool for the designer, it remains a controversial subject among research groups since it is not yet associated with a mechanism through which the failure takes place.

Assuming its correctness, however, the intelligent application of the hypothesis presents serious engineering difficulties. In spite of our present skills in defining the real maximum loads and stresses, one rather serious gap exists in defining the temperature of the blank,  $T_B$ , in the simple equation below:

$$T_T = T_B + T_f$$

where  $T_T$  is the total conjunction temperature and  $T_f$  is the instantaneous computable temperature flash.

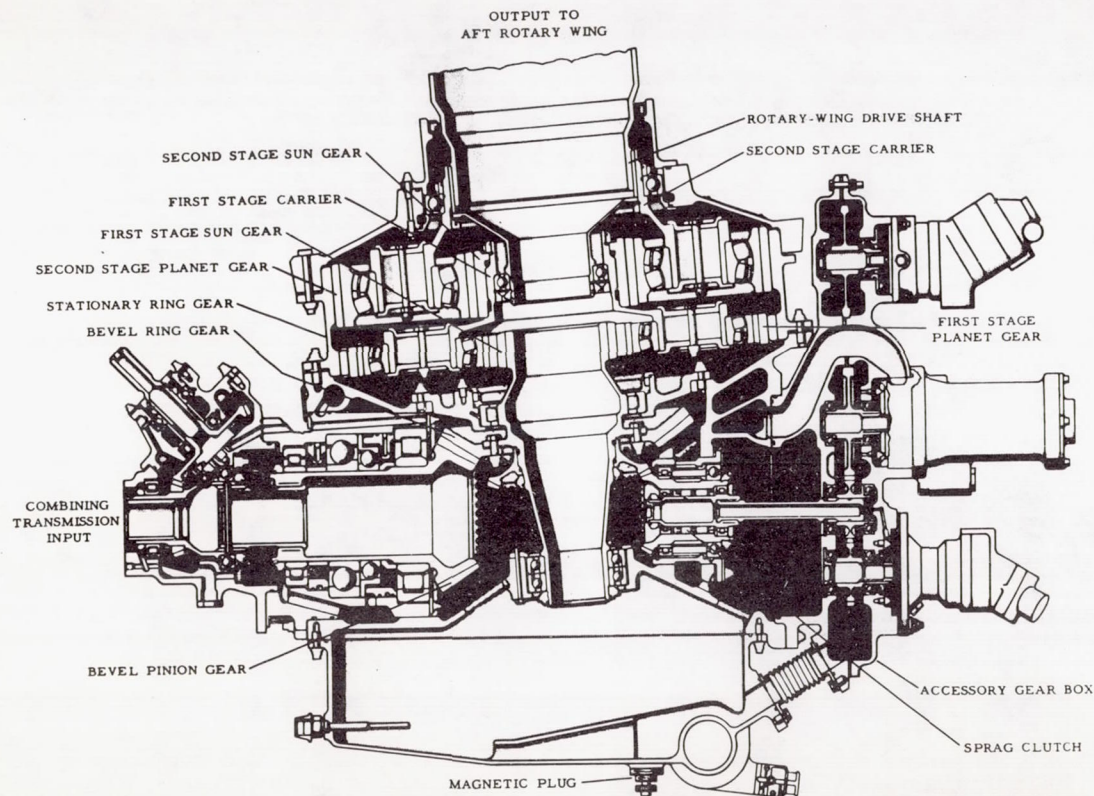


FIGURE 3.—Helicopter transmission.



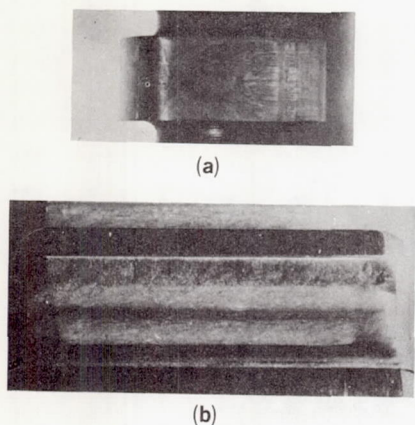


FIGURE 4.—Examples of scoring;  
a—scored cam, b—scored gear.



FIGURE 5.—Scoring of carburized and  
hardened gears in straight mineral oil.

For small high-horsepower components of today, it would be poor judgment to assume that the gear blank temperature is the same as that of the oil. Its importance is even more major than indicated above since the blank temperature controls oil inlet temperature to the conjunction area, thereby controlling friction and the resultant temperature flash. Since this, in turn, controls heat input and blank temperature, the iteration for a complete theoretical solution is obviously not an easy one. The need, however, for something better than that available today is clear.

It has been speculated that "critical temperature" might apply to: (1) the onset of instability of the elastohydrodynamic oil film (ref. 7), (2) the failure of adsorbed films (ref. 8), or (3) melting or other forms of destruction of the contaminants.

Before venturing further into postulates of failure mechanism, it seems important to define, as well as we can, the state of lubrication in concentrated contact machine elements. In this regard, the author finds a strong tendency among researchers and designers to misconstrue some of the implications given in the captivating field of elastohydrodynamic (EHD) lubrication. That this important work has created real insight into the state of lubrication cannot be argued. Professor Dowson, a principal contributor to this relatively new field, will make this quite clear later in the symposium, but the application of this deserves some comment now.

The principles of EHD lubrication have shown that a greater portion



of the load in most cases of concentrated lubricated contact is born by a highly viscous film of oil. Load is not basically supported on the tips of asperities, nor does asperity interaction play as important a role in the overall frictional force as was once thought to occur. EHD lubrication, however, does not preclude metallic contact in most machine elements.

For example, the maximum computable oil film thickness in the transmission of figure 2 is about  $12\ \mu$  in. with the average for the more heavily loaded gears and bearings in the order of  $6\ \mu$  in. The film thickness for elements in figure 3 is not likely to exceed  $10\ \mu$  in. because of the low-viscosity MIL-L-7808 lubricant used. The final drive gears and bearings of large earthmoving vehicles and the bevel gears in trucks and automobiles frequently operate in the  $1$  to  $3\ \mu$  in. range. One is actually hard pressed to find a machine element other than, say, mainshaft bearings on turbines that is capable of operating without some statistical percentage of metallic contact with the finest of commercial finishes on these parts.

However, large quantities of metallic contact with "nonreactive" lubricants do not necessarily result in scoring. In fact, Landen (ref. 9) and others have found that even at computed film thicknesses of less than  $1\ \mu$  in., smooth wear occurs until some critical temperature is reached (more will be said about wear later). An important observation recently made by Earles and Powell led them to conclude in the case of unlubricated steel that "... surface breakdown is unlikely to occur if  $\theta_B$  (the mean surface temperature) ... is less than  $200^\circ\text{C}$ ." This value is not much different from that published for lubricated contact (ref. 10).

These observations leave the field wide open for new postulates, including one that says the critical temperature mechanism pertains to the steel and its oxides and has nothing to do with the lubricant. Certainly such a concept would not conflict with data showing a marked difference in the scoring resistance of gears made from different steels (ref. 11).

As you will note, I have referred to the use of critical temperature as a criterion for failure in the presence of "nonreactive" oil. There is little doubt that "nonreactive" oils in the strictest sense do not exist, but it remains the opinion of the author that the beneficial effects of the reaction products formed and the rate at which they are formed with plain mineral oils are so low as to be of no practical importance in modern machinery.

The real problem is represented by the fact that very few lubricants are now marketed without additives of some form or another, and even fewer will be available in the future—particularly additives that are recognized as being strongly surface-active to prevent or reduce seizure or wear.

The fact that relatively small quantities of the more common sulphur, phosphorous, and chlorine compounds increase scoring resistance signifi-

cantly gives the designer a new plateau. Unfortunately the plateau has not been defined in useful engineering terms.

As a result, industry frequently falls back on full-scale tests and no doubt will continue to do so. But prototypes of the transmissions in figures 2 and 3 cost many hundred thousands of dollars and involve years of test time, and the designer cannot well afford to make a basic mistake in the engineering handling of the lubricant. The problem is far from being a simple one, but the industrial researcher finds it an easier life if he lives in a state of perpetual optimism.

There appear to be at least two mechanisms of failure for extreme pressure (EP) lubricants for which the engineer needs design tools. One of these is the apparent substitution of high rates of wear for scoring. For example, figure 6 shows wear produced by four such oils on a disc type machine (ref. 12), two of them showing excessive wear. Unfortunately such wear cannot be tolerated on gears or engine cams because it occurs unevenly on the profile leading to other modes of failure such as pitting (ref. 5). One of the first problems in exploring this field is how to obtain significant data in a reasonable length of time without the costly use of equipment such as radioactive tracers.

One possibility seems to be in the development of some understanding of the frictional response with EP additives. For example, with a common

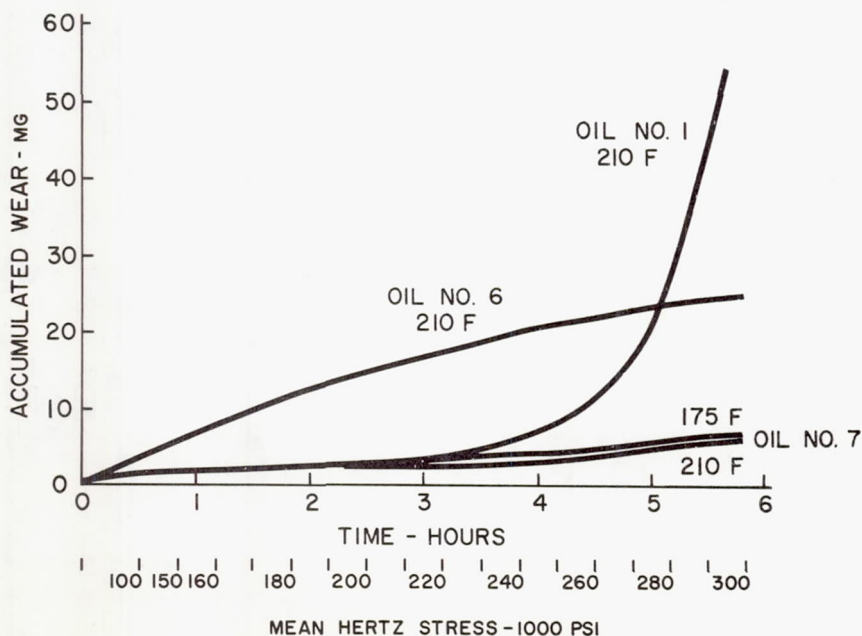


FIGURE 6.—Wear rates with EP oils.

sulphur, phosphorous, or chlorine additive, the friction on a disc test machine changes as the load is increased (fig. 7); at some load and computed conjunction temperature,  $T_c$ , the coefficient of friction starts to drop, holding the temperature constant over a sizable load range. Finally friction levels out, allowing temperature to rise until failure occurs by scoring. It is the opinion of the author that such measurements are meaningful and will ultimately correlate with the state of lubrication and the onset of EP wear.

The other EP mechanism is that of the development of strong durable films formed by additives such as zinc dialkyl dithiophosphate. Wear does not appear to occur with these, and the author is not aware of published work on the mechanism for its failure.

Two other related quantifying areas are needed for the engineer. One is the reaction rate characteristics of the EP additive since it can be shown that, by increasing the rate of speed change at a fixed load on gears or in disc machines, the formation rate of protective films can be exceeded such as is shown in figure 8. The second is the reaction as a function of gear blank temperature. Whyte (ref. 13), for example, has shown that under cold ambient temperatures, EP additives do not respond. As shown in figure 9, the 75-grade lubricant with 10-percent additive concentration left a "temperature gap" where scoring protection from viscosity failed before gear blank temperatures that would produce

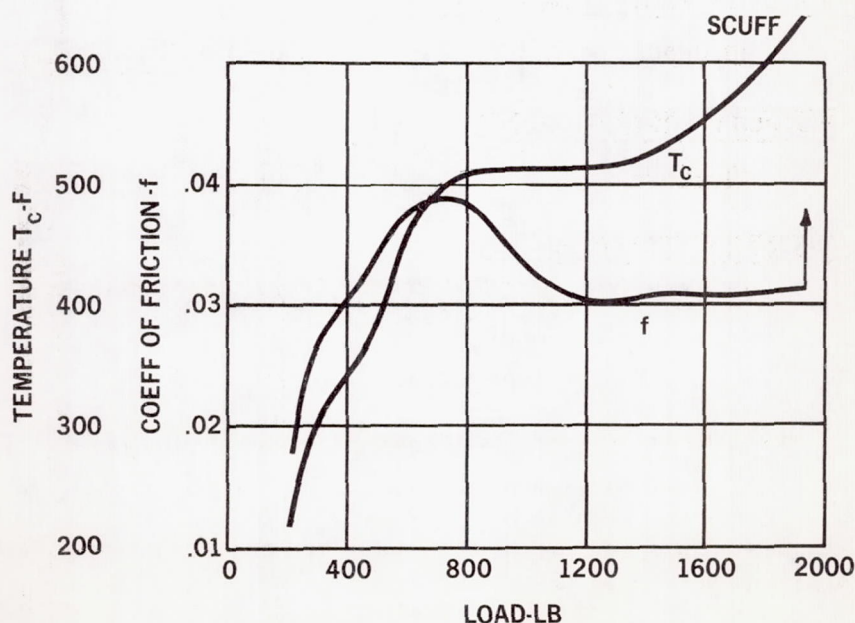


FIGURE 7.—Friction response of an EP oil.



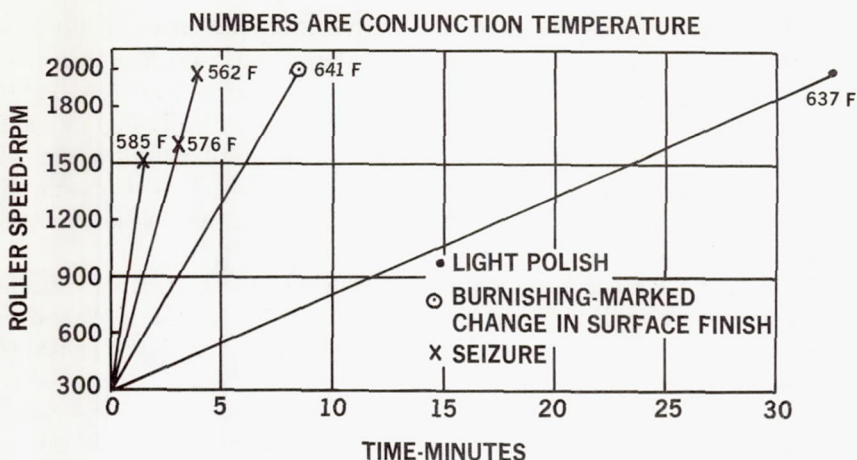


FIGURE 8.—Reaction rate results with an EP oil.

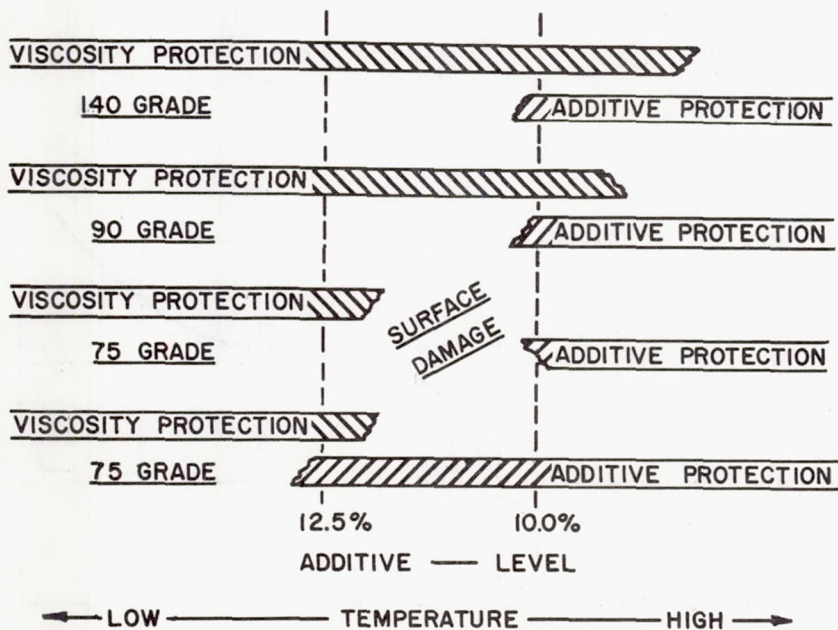


FIGURE 9.—Temperature gap of EP oils.

adequate additive reaction were reached. This can also be shown on disc type machines (ref. 5).

#### WEAR

For many years the author sincerely felt that the lubrication of counter-formal elements was principally a loosely termed "boundary phenome-

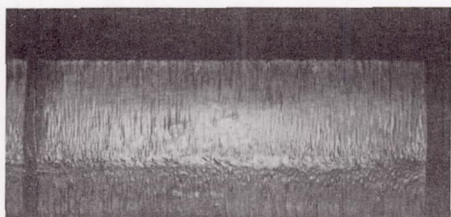
non." Yet it was remarkable indeed that the wear rate of such parts under most conditions was so low that one frequently found finish machine marks still evident after many years of service. This apparent conflict has been impressively explained by the work on EHD lubrication.

In the writer's opinion, one of the most revealing experimental papers on the subject of EHD and wear was published by Landen (ref. 9), a paper worthy of interpreting here in some detail in terms of the practical problem. First of all, the careful observer can find thin-film wear occurring on low-speed final-drive gears, hypoid gears, and small fuel-pump gears as shown in figure 10. To the untrained eye, it may have an appearance similar to scoring or scuffing; the difference lies in the generally softer, untorn texture and in the fact that on gears it frequently occurs all the way to the pitch line, thus showing a general independence of wear to sliding velocity.

Using carburized and hardened rollers, Landen has shown that such



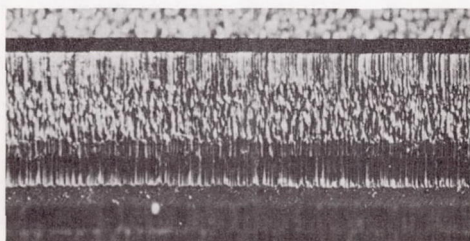
**HYPOID GEAR**



**SPUR GEAR**

**TOOTH TIP**

— **PITCH LINE**



**FUEL PUMP GEAR**

**TOOTH TIP**

— **PITCH LINE**

**FIGURE 10.—Examples of thin-film wear.**



continuous forms of wear can be predicted when the computed oil-film thickness is on the order of  $1.5 \mu$  in. or less as shown in figure 11. Surprisingly the results also showed that rippling developed in the transition between continuous and no wear. Figure 12 shows that this form of surface damage can be found on final-drive gears and bevel gears. Such rippling was described previously as being the result of severe plastic flow of the material (ref. 14). Closer observations have shown that while plastic flow may be present, rippling may occur without the onset of plastic flow. Rippling, therefore, is a form of true wear. Interestingly

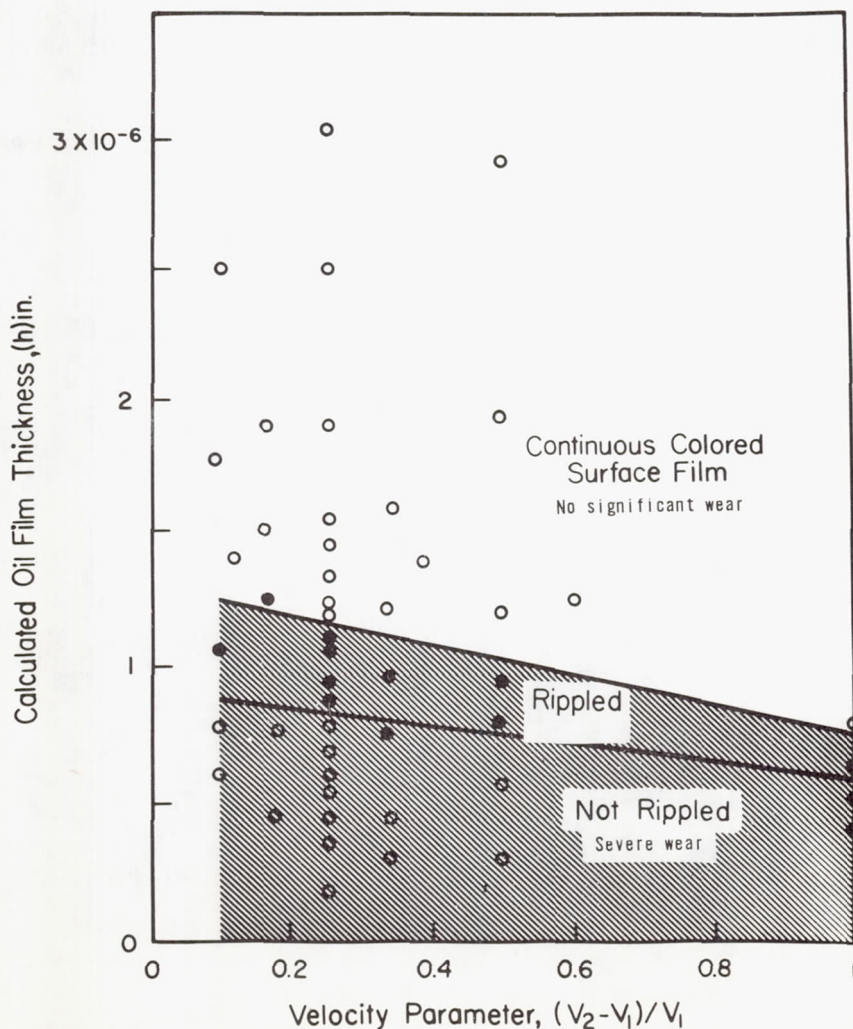
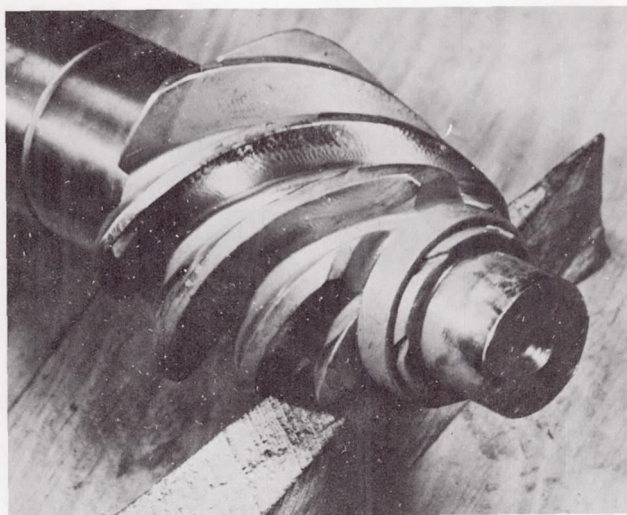
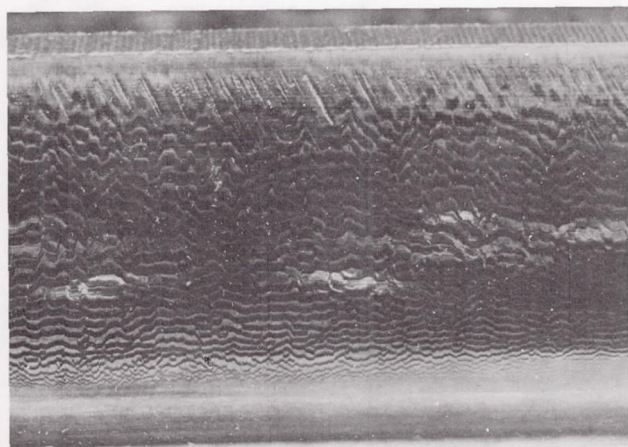


FIGURE 11.—Control of wear by film thickness.

**HYPOID GEAR****TOOTH TIP****PITCH LINE****LOW SPEED SPUR GEAR****FIGURE 12.—Examples of rippling.**

rippling occurs only on the element having the faster velocity, such as the addendum of gears, and has been found to have an unexplained peak-to-peak frequency related to velocity parameter  $(V_2 - V_1)/V_1$ .

It is also noteworthy in figure 11 that continuous color films from oil oxidation products form in the no-wear regime and in the valleys of the ripples. One can interpret this as an indication that such films help prevent wear by their properties, or that they are permitted to form as a

result of an adequate elastohydrodynamic film thickness. Crude exploration of the film's physical properties by scratching with a penknife indicates that the latter case is more logical.

The work mentioned above was performed with a straight commercial grade of mineral oil falling into the writer's category as "nonreactive." In more recent unpublished work, Landen has found that continuous wear can be subdued with EP additives. For example, whereas computed oil-film thickness of  $1\text{ }\mu\text{ in.}$  will create continuous wear with "nonreactive" oils, a common sulphur, phosphorous, or chlorine additive will allow a computed film thickness of less than  $0.3\text{ }\mu\text{ in.}$  without significant wear. An interesting transient occurs, as is shown in figure 13, wherein the initial wear rate with the EP additive was markedly greater than that of the "nonreactive" oil, but within a couple of hours leveled off to a very low rate.

#### CONTACT FATIGUE

Contact fatigue or pitting has long been a limiting parameter for the designer working on elements having concentrated contacts. The generally used criterion for defining the onset of pitting failure is contact stress (Hertz). Only small gains can be made in reducing these stresses in either gears or bearings by geometric change within a given size. For example, the diametral pitch of gears can be coarsened to eliminate a fracture problem, and EP additives can be used to upgrade seizure re-

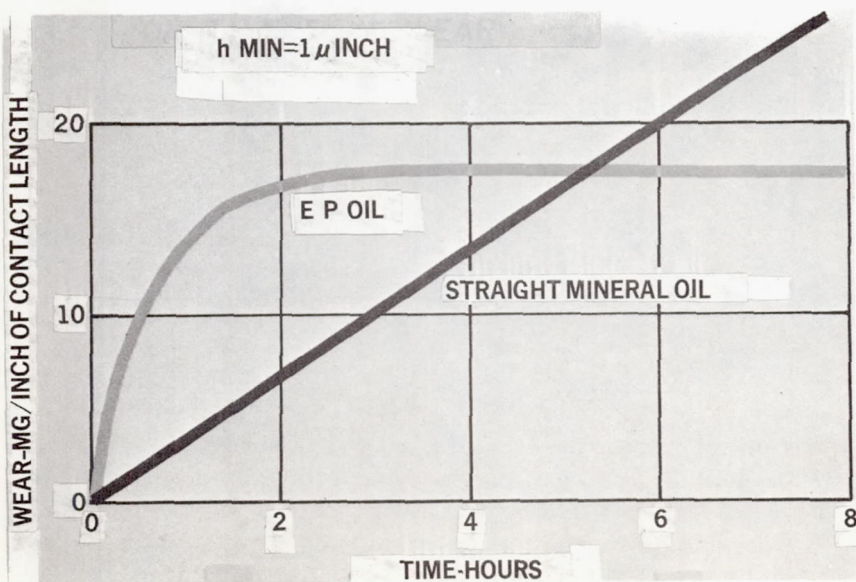


FIGURE 13.—Effect of EP additive on thin-film wear.



sistance; but only marginal reduction of contact stresses can be achieved at a given center distance and face width.

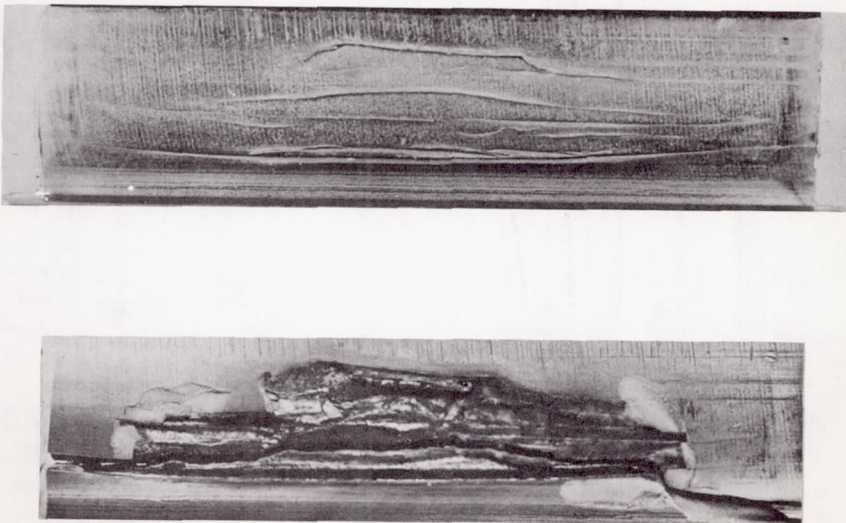
Sizeable increases in contact fatigue resistance have been made over the years with improved materials and heat treatment, but the rate of improvement tends to decrease for those who make the most of such metallurgical advances.

Fortunately the future does not look so hopeless when one considers the effect of lubricant viscosity and chemistry on pitting. In fact, a serious look at such environmental factors of practical machine elements leads one to feel that we are far from achieving the optimum performance level of even modest quality steels.

Dr. Littmann will, I am sure, discuss thoroughly the competitive modes of contact fatigue for rolling-element bearings; but the writer, representing a user of bearings and gears, would like to have some biased impressions as to their importance.

One type of contact fatigue (fig. 14) known as case crushing (or sometimes "sub-case fatigue") can be dispensed with quickly. It is strictly a stress-to-strength problem associated with the stresses and metallurgical characteristics at the junction of case and core of case-hardened parts (ref. 15). The designer has little excuse for being trapped by the problem, and there appears to be no relationship between the mode of failure and environmental factors.

The classic mode of contact fatigue failure is that caused by shear



### FINAL CASE CRUSHING FAILURE

FIGURE 14.—Example of case crushing failure.



stresses,  $\tau_{45}$ , below the surface due to the nonhydrostatic character of Hertz compressive stresses as shown in figure 15. Many stress analysts and researchers have spent months and even years exploring the stress field and the materials in the areas of high shear stresses in hopes of solving the problem or defining the cause of contact fatigue. It is the writer's opinion that such a "classic" mode of failure in reasonably clean material is so rare as to be of little practical importance in the great majority of applications.

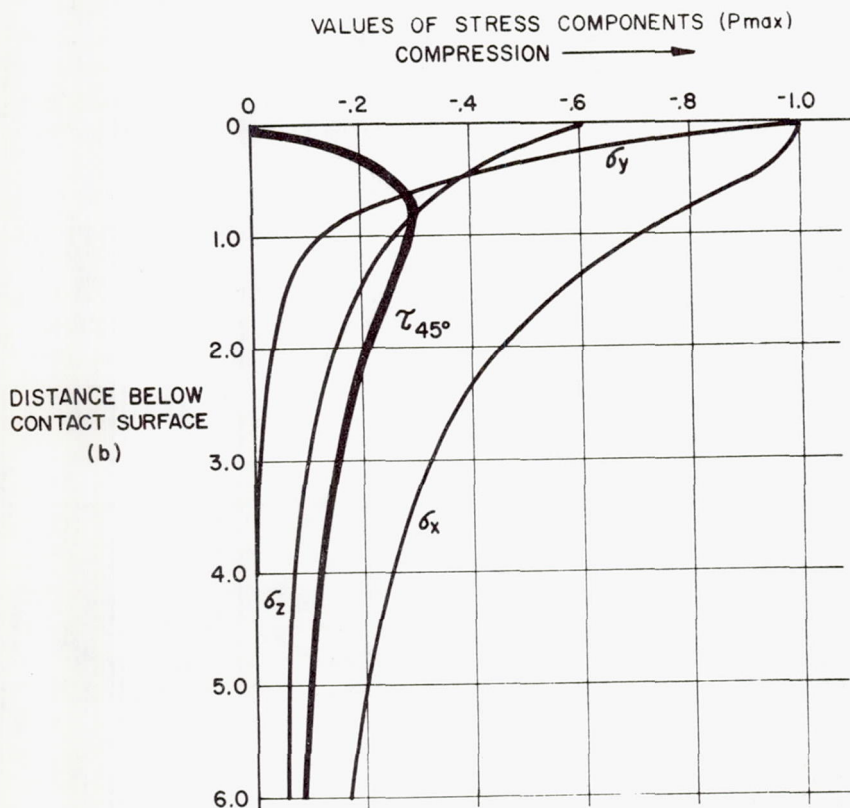


FIGURE 15.—Distribution of stresses in a cylinder due to a normal load.

The definition of the real problem is a more difficult one, and perhaps a part of it is best illustrated by the chart in figure 16 by Tallian (ref. 16). It shows that, as the ratio of minimum EHD film thickness to composite surface roughness increases, the contact fatigue life is increased very significantly—in the order of perhaps 40 times (ref. 17). In well-controlled laboratory tests with thick EHD films and no sliding, capacities beyond 600 000- to 800 000-psi maximum Hertz stress are necessary to create

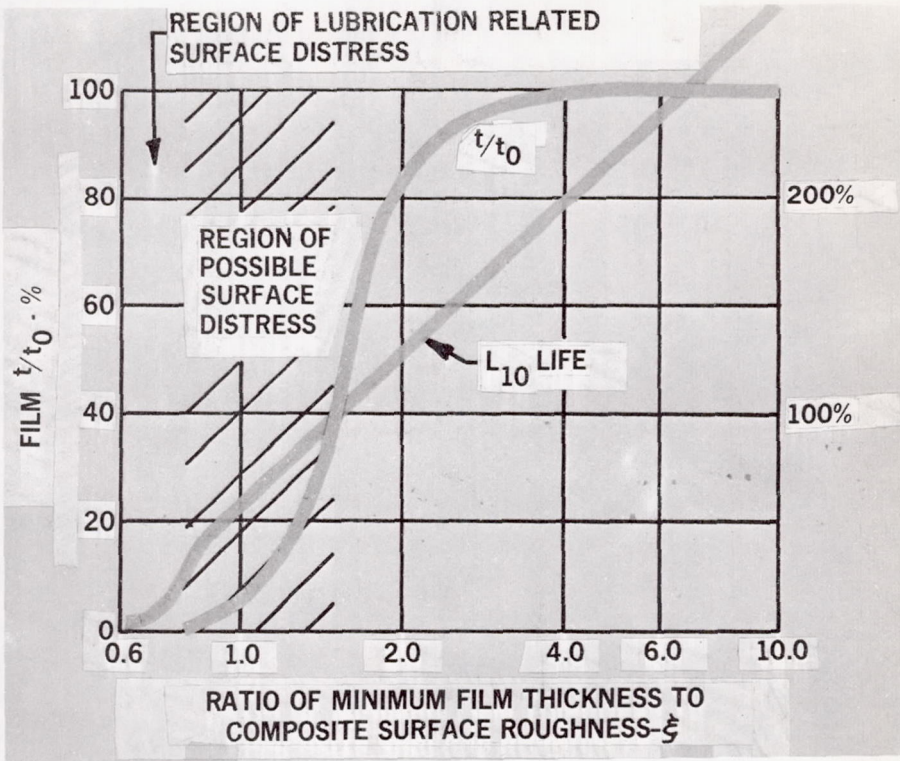


FIGURE 16.—Effect of film thickness on bearing life.

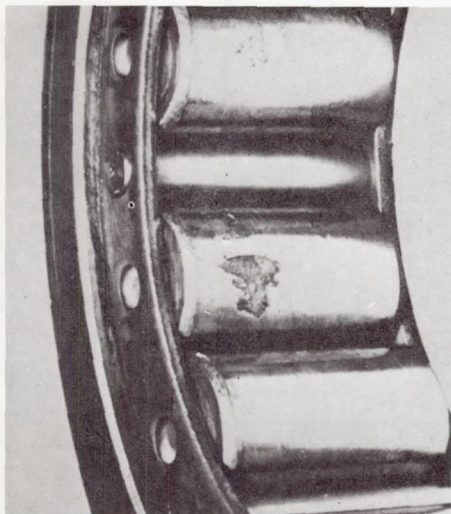
failure in any reasonable length of time, and it has been the writer's observation that these are the classic subsurface types. If rolling-element bearings, gears, or engine cams could operate at such stresses, we would be fairly well out of the contact fatigue problem.

Considering, however, previous comments on the rarity of complete EHD separation in the great percentage of machine elements, the ill-defined problem lies on the lower end of the curve which says "region of possible surface distress." (The writer hastens to add that some of the infrequent applications where full EHD separation can be achieved, such as high-speed gas turbine bearings, are important ones, and personal discussions with engineers responsible for the reliability of such bearings have indicated that the application of EHD principles has led to the virtual elimination of contact fatigue in those cases.)

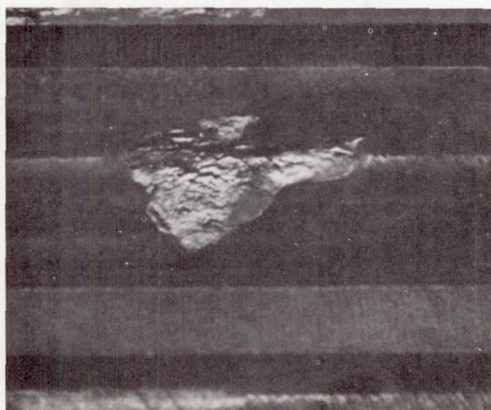
When we deal with this more general problem of the persistent metallic contact, several observations from practical experience are important. First is the general recognition that contact fatigue in such instances is initiated at or very close to the surface. Two modes of pitting illustrate this on gears and bearings. One is the typical arrowhead form of pit

shown in figures 17 and 18; the other is a mode of failure referred to as frosting or superficial fatigue as shown in figure 19.

At this point frosting is perhaps the most mysterious. Since it occurs only on elements under modest to low oil-film thicknesses, it is obviously affected by the existence of metallic contact. An important observation, however, is that it does not always occur under these conditions. Changing a batch of bearings or a heat of steel frequently results in elimination of



**ROLLER BEARING**



**TOOTH TIP**

**PITCH LINE**

**START OF  
ACTIVE PROFILE**

**GEAR**

FIGURE 17.—Typical surface-initiated pitting.



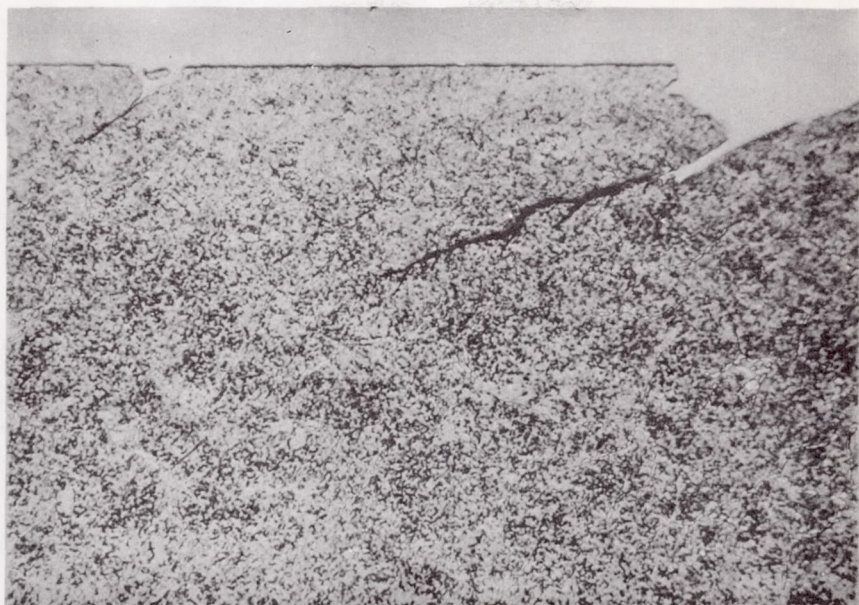
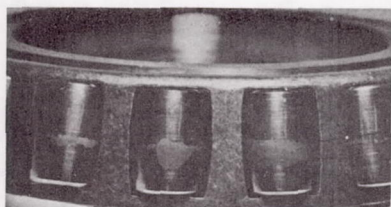


FIGURE 18.—Photomicrograph of surface-initiated pitting.



ROLLING ELEMENT BEARING



GEAR



MAGNIFIED VIEW OF SURFACE

FIGURE 19.—Superficial fatigue.



the appearance in any series of tests. This response indicates that it may also be affected by material and/or possibly machining method.

Researchers do not agree that frosting leads to more serious pitting, but I suspect that this is partly due to our natural reluctance to pay attention to something we do not understand. It is our experience that the appearance of frosting means that full-scale pitting failure is imminent.

It is generally recognized that contact fatigue is sensitive to the type and amplitude of sliding as well as to the state of lubrication. Positive sliding is defined as that which occurs on the faster of two elements in contact. This could occur, for example, on the addendum area of a gear tooth (that area above the pitch line). Negative sliding occurs on the mating surface, for example, the dedendum of gears. Peculiar things happen on the negative sliding surface. Not only do they pit more readily, but they also wear much faster than can rationally be explained and they do not ripple. (Rippling occurs only on the positive sliding surface.)

Explanations as to why the element in negative sliding should pit more readily are still far from satisfactory. An examination of the thermally induced stresses shows somewhat greater compressive stresses on the surface of the negative sliding element (ref. 18), but failure-creating shear stresses are only marginally greater. Way's explanation (ref. 19), which considers the trapped oil in an existing crack as a form of propagation in the negative sliding element, still does not explain why even incipient cracks do not form on the positive sliding element (ref. 20).

One can say with little danger of rebuttal that pure rolling of machine elements (except at the operating pitch line of gears) only occurs in the laboratory under very sophisticated conditions. The slide-to-roll ratio on the profiles of gears can easily be determined. Rolling-element bearings, while less defined, are also subjected to sliding as well as rolling because of the frictional resistance of the cage and rollers. The slippage in such bearings may be only 1 or 2 percent, whereas that occurring on gears such as hypoids may be several hundred percent, but there are indications in the literature that a few percent can be serious—particularly with harder materials.

Unfortunately the well-documented evidence on the effect of sliding on contact fatigue has been with the use of mild steel. One of the most frequently quoted articles by Dawson (ref. 20) shows that with a plain carbon normalized steel (of about 4.60 mm Br), pitting fatigue in negative sliding is not greatly affected by slide-to-sweep ratio as shown in figure 20. This same result can be extracted from the classic work of Talbourdet (ref. 21). The chart in figure 21 is constructed from his data (at  $100 \times 10^6$  cycles) which show that lower-hardness materials are not overly sensitive to the percent of sliding. Only two points are published by Talbourdet for Rc 58 steel, and the writer has taken the unethical liberty of drawing a curve through them which seems reasonable. This

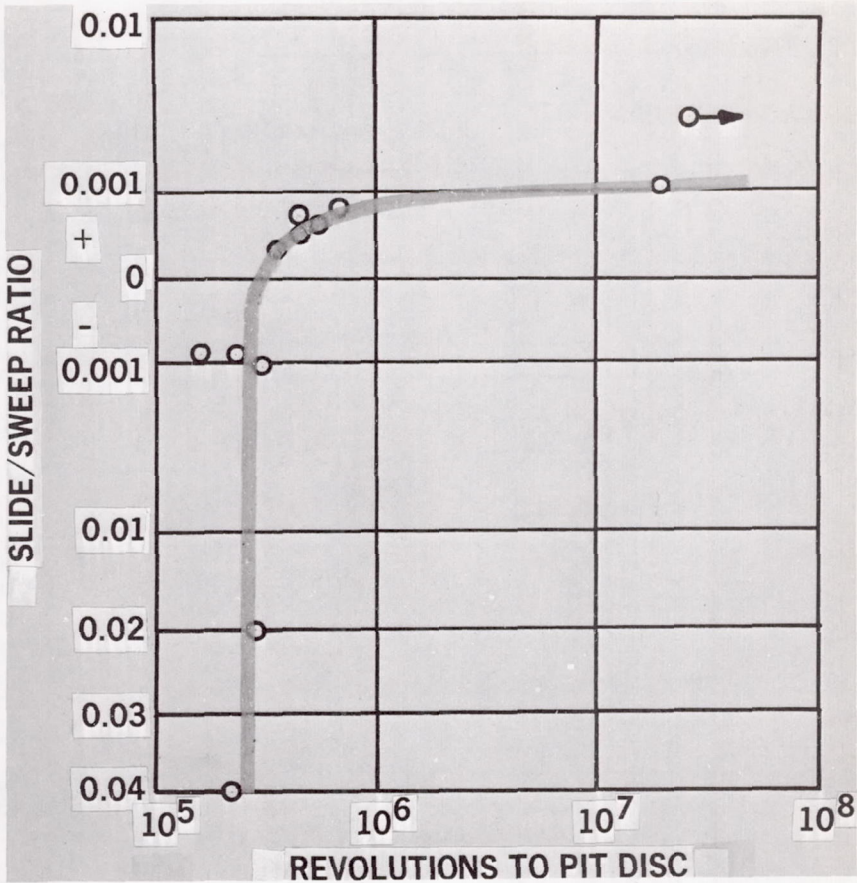


FIGURE 20.—Effect of sliding on pitting with mild steel.

indicates that even a small percentage of sliding is capable of reducing load-carrying capacity rather significantly on the harder materials. This effect is consistent with observations of pitting failure of soft steel gears and carburized and hardened gears. As shown in figure 22, the softer material pits at random in the negative sliding area even though the sliding increases linearly with distance from the pitch line, whereas the initiation of pits on carburized and hardened gears are generally far from the pitch line (fig. 17). Incidentally so-called "pitch line pitting," frequently mentioned in the literature, is largely due to inaccurate observation or the nonuniform destruction of the profile caused by wear as a preliminary mode of failure.

A variable of great practical importance has added further to the complex problem on contact fatigue. This is the chemistry of the lubricant.



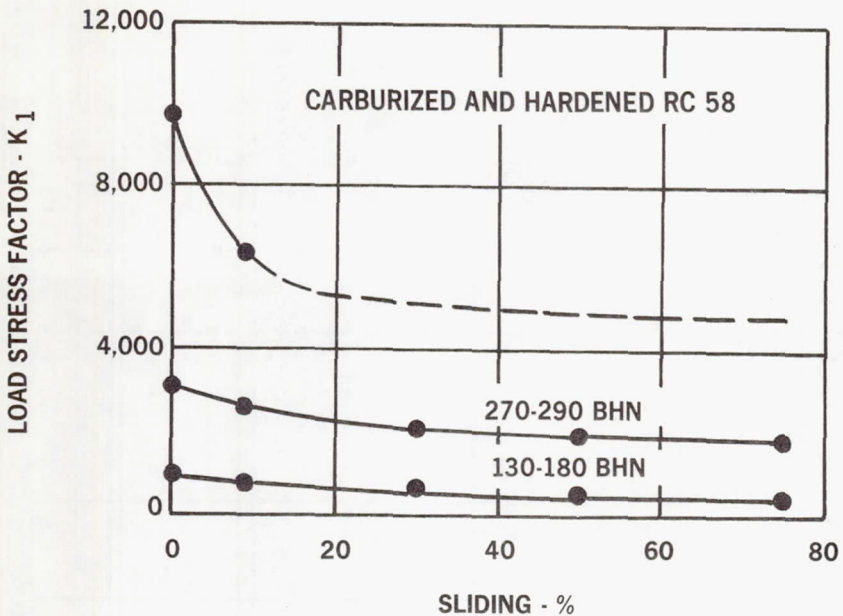


FIGURE 21.—Effect of sliding on pitting load (lb/in of face width) for  $100 \times 10^6$  cycles (typical data from work of Talbourdet).

Using roller test machines at a maximum Hertz stress of 380 000 psi with combined sliding and rolling (surface velocities 75 and 52 ips), we have attempted at least to quantify the danger as illustrated in figure 23. Two rather alarming characteristics of the Weibull distributions show up in these tests. First is that some EP additives readily available to customers in the field can reduce contact fatigue life several-fold. Second is the observation that the slope of the Weibull curve is steeper when chemical effects are present. While this latter characteristic, indicating a lessening of the scatter of results, is enjoyed by the engineer running the test, it leaves the parts manufacturer a little nervous because of the increased number of failures if the reduced life is exceeded.

The amount of evidence showing detrimental chemical attack from lubricants will probably increase in the future, and you are likely to hear of it again in this conference. One disheartening find by Scott and Blackwell (ref. 23) is a reversal of comparative performance rating with one steel as compared to another. For example, one gas turbine lubricant shows a significantly lower life using an M10 steel as compared to an En 31 steel. The reverse is true with another lubricant. This indicates the potential danger to manufacturers who develop new materials or better heat treatments for gears or rolling-element bearings. The performance of the new metallurgy could be worse rather than better with a



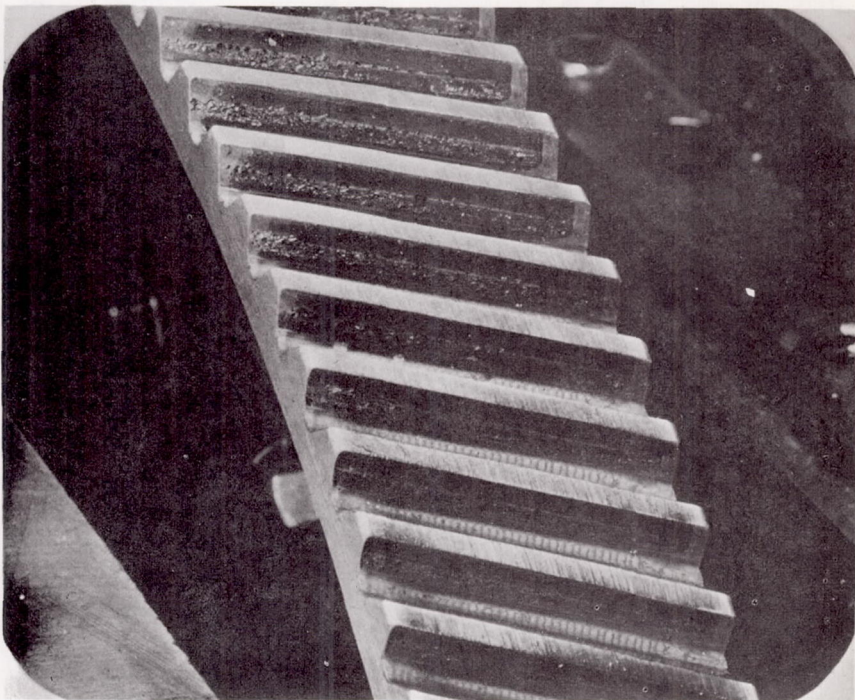


FIGURE 22.—Pitting of machinable-hardness gear.

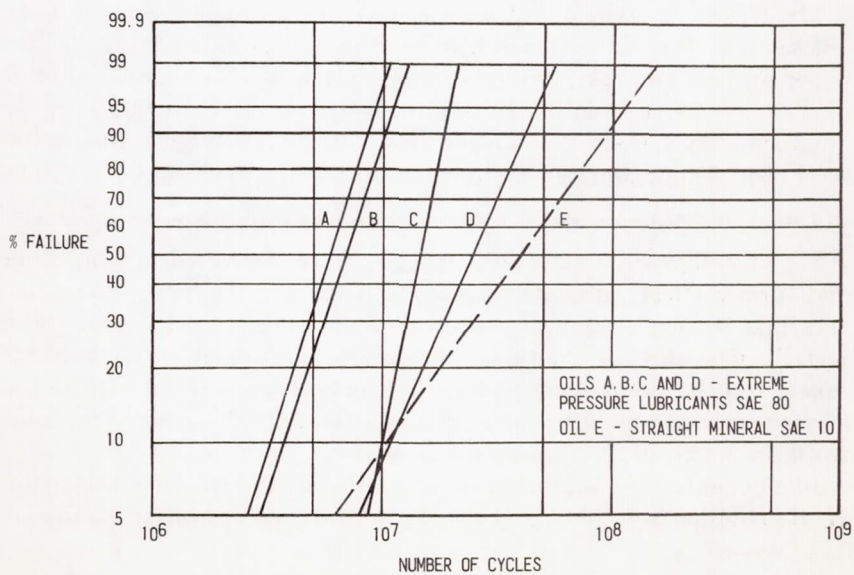


FIGURE 23.—Effect of some EP lubricants on pitting fatigue life.

different lubricant. Needless to say, the multiplication of tests to screen such variables becomes ridiculously high. The only way is through understanding of the mechanism that is at work.

Finally a remark about tests and test equipment. The writer has been closely associated with the extraction of useful information from test equipment, such as gear tests and roller tests, for over 20 years. Gear test machines are useful in the understanding of factors such as geometric effects and the kinematics of gear problems as long as they are handled intelligently. The complexity of a pair of gears, much less a full transmission, however, is sufficiently filled with variables that one is hard pressed to bring forth any useful or universal design criteria on the problems of concentrated contacts. We have probably learned more about gears by trying to make rollers act like gears in terms of contact fatigue, scoring, and wear than by any other method of testing. In general we reached the conclusion many years ago that was recently reached by Manton and O'Donoghue (ref. 23); i.e., once the researcher is reasonably familiar with the basic engineering principles of such elements, he is well advised to stay away from gear test machines, such as four-square machines, except in very special cases. Such units that are reinvented almost annually tend only to perpetuate the lack of any real understanding in the problems of concentrated contact.

#### DISCUSSIONS

##### **H. Klein (Lycoming Division, AVCO Corporation, Stratford, Connecticut)**

Mr. Kelley mentioned that we can have metal-to-metal contact without scoring. The converse can also be true. As shown in figure 13 of a paper by Brix (ref. 24), scoring can take place on a case-hardened steel disk under a brass pin at a distance of approximately 0.001 inch, at approximately 1200-ft/min linear speed. The pin produced a score in the steel disk of approximately 0.001-inch width and a depth of 0.0001 inch.

##### **A. J. Gentile (AC Electronics, Division of General Motors Corporation, Milwaukee, Wisconsin)\***

For several years, I have been engaged in work devoted to lubrication and surface effects in contact loaded applications. There is, I feel, one area that is completely neglected in these investigations and that is the metallurgical effect of the surface preparation. Although extreme care is taken to reproduce specimens, there is no effort made to investigate the surface condition on either a sample or a 100-percent basis. As a result, the data can be off by a significant amount.

In a paper being co-authored by D. McCormick of the New Departure Hyatt Division of G.M.C. and myself, we will present bearing endurance

---

\* See also Dr. Littmann's closure at the end of his lecture entitled "The Mechanism of Contact Fatigue."



results illustrating this point. In these tests, we varied the surface preparation on a group of bearings to include various mechanical and chemical finishing techniques and found a 2:1 difference in the B-10 life between the lots with the same lubricant. The number of bearings involved was sufficient to lend statistical significance to these results. Metallurgical examinations reveal considerable difference in surface properties depending upon finishing treatments utilized.

These results emphasize the point that any testing involving one metal in contact with another, with or without lubrication, should definitely include a metallurgical assessment of the surface.

#### LECTURER'S CLOSURE

The writer can do no more than agree completely with the comments by Mr. Gentile concerning the negligence of investigators dealing with lubricant failures in the area of steels whose chemistries and metallurgical characteristics are not well defined. While I am sure that Mr. Gentile has concerned himself primarily with the area of contact fatigue, I would like to emphasize that the metallurgical properties of the materials are equally important when it comes to seizure or scoring. The term "scoring" is more frequently used to define the tearing of surfaces than other damaging effects such as severe wear or smearing. The resistance of any material to scoring as defined above is markedly affected by the chemistry of the material and the metallurgical structures that existed prior to this mode of failure. On such bench tests as the four-ball machine, for example, the writer believes that, before actual seizure, severe metallurgical changes take place that may very well be sensitive to the chemistry, the amount of retained austenite, intermediate transformation products, and the load history prior to reaching the failure point.

The writer can only re-emphasize Mr. Gentile's comments concerning the real need for those who are testing lubricants to become familiar with the properties of the materials.

*Editor's note*—Mr. Kelley has elected not to respond to Dr. Klein's reference to Brix's work. The editor has read Brix's article several times, but fails to understand the experimental evidence and conclusions advanced therein.

#### REFERENCES

1. DAVISON, C. ST. C. B.: Geared Power Transmission. The Chartered Mechanical Engineer, March 1962.
2. USHER, A. P.: A History of Mechanical Inventions. Harvard Univ. Press, 1954.
3. BROWN, E. D., JR.: Medical Tracts, Patent 543, General Electric Co., Albany, N.Y.
4. ANON.: Aft transmission of the CH-47 Chinook, courtesy of the Vertol Division, The Boeing Company.



5. KELLEY, B. W.; AND LEMANSKI, A. J.: Lubrication of Involute Gearing. *Proc. Inst. Mech. Engrs.*, vol. 182, pt. 3A, 1967-68, p. 173.
6. BLOK, H.: Surface Temperatures Under Extreme Pressure Conditions. *Proc. 2nd World Petroleum Congress (Paris)*, 1937.
7. IBRAHIM, M.; AND CAMERON, A.: Oil Film Thickness and the Mechanism of Scuffing in Gear Teeth. *Proc. Conf. Lubrication and Wear, Inst. Mech. Engrs.*, 1963, p. 228.
8. ASKWITH, T. C.; CROUCH, R. F.; AND CAMERON, A.: The Relationship of Lubricants and Additives for Optimum Effect; and a Theory of Scuffing. *Proc. Sym. Gear Lubrication, Inst. of Pet. (London)*, 1966.
9. LANDEN, E. W.: Slow Speed Wear of Steel Surfaces Lubricated by Thin Oil Films. *ASLE Trans.*, vol. 11, 1968, p. 6.
10. EARLES, S. W. E.; AND POWELL, D. G.: Surface Temperature and Its Relation to Periodic Changes in Sliding Conditions Between Unlubricated Steel Surfaces. *ASLE Trans.*, vol. 11, 1968, p. 109.
11. BABER, B. B.; ANDERSON, E. L.; AND KU, P. M.: Effects of Lubricants, Metals, Temperature, and Atmospheric Environments on Gear Load-Carrying Capacity. *Trans. ASME, J. Lub. Tech.*, vol. 90F, 1968, p. 117.
12. TAYLOR, A. E.; AND WILDE, R. A.: New Test Techniques for Evaluating Gear Oils. *Lub. Engrg.*, vol. 20, 1964.
13. WHYTE, R. B.: Laboratory and Field Evaluation of SAE 75 Grade Gear Oils. *Proc. Sym. Gear Lubrication, Inst. of Pet. (London)*, 1966.
14. COLEMAN, W.: Bevel and Hypoid Gear Surface Durability: Pitting and Scuffing. *Proc. Inst. Mech. Engrs.*, vol. 182, pt. 2A, 1967-68, p. 191.
15. PEDERSEN, R.; AND RICE, S. L.: Case Crushing of Carburized and Hardened Gears. *SAE Trans.*, 1961, p. 370.
16. TALLIAN, T. E.: Rolling Contact Failure Control Through Lubrication. *Proc. Inst. Mech. Engrs.*, vol. 182, pt. 3A, 1967-68, p. 205.
17. TALLIAN, T. E.: On Competing Failure Modes in Rolling Contact. *ASLE Trans.*, vol. 10, 1967, p. 418.
18. KELLEY, B. W.: The Importance of Surface Temperature to Surface Damage. *Handbook of Mechanical Wear*, Lipson and Colwell, eds., Michigan Univ. Press, 1961.
19. WAY, S.: Pitting Due to Rolling Contact. *J. Appl. Mech.*, vol. 2, 1935, p. A49.
20. DAWSON, P. H.: The Pitting of Lubricated Rolling Surfaces. *Power Transmission*, April and May, 1961.
21. TALBOURDET, G. J.: Surface Endurance Limits of Various USMC Engineering Materials. *Rept. of the Research Division, United Shoe Machinery Corp.*, 1957.
22. SCOTT, D.; AND BLACKWELL, J.: Study of the Effects of Elevated-Temperature Lubricants on Materials for Rolling Elements. *Proc. Lubrication and Wear, 5th Conv., Inst. Mech. Engrs.*, vol. 181, pt. 30, 1967-68.
23. MANTON, S. M.; AND O'DONOGHUE, J. P.: Gear Lubricant Testing Machines—Are They Useful? *Tribology*, Aug. 1968.
24. BRIX, V. H.: Scoring and Burnishing in Bearings. *Aircraft Engrg.*, vol. 19, no. 221, July 1947.

# Elastohydrodynamic Lubrication

**D. DOWSON**

**The University of Leeds  
Leeds, England**

The history of this relatively new topic in tribology is presented, and attention is drawn to the significant theoretical and experimental developments. The position of elastohydrodynamics in the spectrum of lubrication regimes is discussed, and the essential features of elastohydrodynamic contacts are described. Attention is focussed upon film thickness calculations and measurements since this represents one of the most important and best charted fields within the subject, but friction and the distribution of pressure are also considered. Transition from hydrodynamic to elastohydrodynamic and boundary lubrication conditions is considered, and attention is drawn to the importance of the ratio of film thickness/surface roughness. A number of successful engineering applications of elastohydrodynamics are noted, and the present trends in research are outlined.

**T**HE SUBJECT OF ELASTOHYDRODYNAMIC LUBRICATION has arrived and established a place in tribology within the last 15 or 20 years, and this is a good time to take stock of the position in this relatively new field. The subject is well known to many workers concerned with the lubrication of highly stressed machine elements, but there is a danger that the concept may be adopted as a panacea for all situations without due regard to its limitations. The validity of present understanding of the subject and the extent to which it has contributed to knowledge of the science and technology of tribology will be considered. Some of the gaps in our knowledge will be noted, and present research trends will be discussed. The author will also speculate on future developments.

## **HISTORY**

The interesting history of studies in elastohydrodynamic lubrication has been discussed in detail elsewhere (refs. 1 to 3), but it is useful to establish a historical background for the present discussion. Much of the early interest in the subject was generated by the need to understand the mechanism of gear lubrication. Martin's analysis (ref. 4) in 1916 of the possibility of hydrodynamic action in involute gears provides a useful



starting point for an account of the subject. This analysis was based on the assumption that the gears could be treated as rigid solids and the lubricant as an isoviscous fluid. The result of the theoretical study was not encouraging since it predicted film thicknesses that were too small to offer any support for the hydrodynamic lubrication hypothesis.

After an interval of about 20 years, the theoretical analysis of highly stressed lubricated machine elements was reconsidered, and the limitation imposed by the rigid solid assumption adopted by Martin was relaxed. Peppler (refs. 5 and 6) and Meldahl (ref. 7) considered the lubrication of elastic solids by an isoviscous lubricant in the late 1930's and early 1940's; and, although the investigations failed to demonstrate the full significance of elastic effects, they did point the way for future investigators. It was shown that the inclusion of elastic deformation in the analysis improved the predictions of film thickness, but the resulting values were still very small compared with the surface roughness of gear teeth.

During the late 1940's and early 1950's, the importance of the isoviscous lubricant assumption adopted by Martin was examined theoretically by a number of workers. Gatecombe (ref. 8), Blok (ref. 9), and McEwen (ref. 10) considered different pressure-viscosity relationships, and they all predicted a maximum increase in minimum film thickness of about 150 percent. The improvement was similar to that predicted by studies of the influence of elastic distortion and in itself inadequate from the point of view of supporting the hydrodynamic lubrication hypothesis for gear lubrication.

It was clear from the investigations mentioned above that the separate effects of elastic distortion and the increase of viscosity with pressure did not improve the film thickness predictions by an amount that provided any confidence in the possibility of hydrodynamic lubrication in gears. This evidence was at variance with the observed performance of gear sets which suggested that the teeth were protected by a coherent film of lubricant under many conditions. Similar observations had been made on various forms of the disc machine introduced by Merritt (ref. 11) in 1935 to simulate gear contacts.

One of the main reasons for the long delay in taking account of the effects of high film pressures on local elastic distortion and lubricant viscosity simultaneously was no doubt the complexity of the theoretical problem. However, Grubin (ref. 12) provided a definitive answer to the problem when he presented an approximate analytical solution that avoided the necessity to solve the full lubrication and elasticity equations simultaneously. Grubin argued that the shape of the elastically deformed surfaces would be little affected by the presence of an oil film since the latter was likely to be very thin compared with the local elastic compression. He therefore assumed that the shape of the gap between the cylinders



would be given by the normal Hertzian shape and proceeded to integrate the Reynolds equation to determine the separation of the Hertzian "flats" which would allow the pressure to rise to infinity at inlet to the Hertzian zone. Allowance was made for the variation of viscosity with pressure and the calculated gap,  $h_0$ , between the flats known as the Grubin film thickness. His solutions provided a spectacular increase in film thickness above the predictions of classical hydrodynamic theory, and a new appreciation of hydrodynamic action was developed for conditions previously assumed to be governed by surface or boundary lubrication actions. The fact that interactions between elastic distortion and the increase in lubricant viscosity with pressure can have an effect so very much greater than the separate actions is a vivid demonstration that multivariable problems cannot always be tackled safely by considering the role of each variable in isolation.

Grubin's analysis predicted film thicknesses one or two orders of magnitude greater than the classical hydrodynamic lubrication theory for typical gear and roller bearing operating conditions. The analysis did not, however, give details of the film shape and pressure distribution throughout the contact zone. A limited number of more complete solutions based upon numerical analysis of the governing equations was provided by Petrusevich (ref. 13) in 1951.

Dowson and Higginson (ref. 14) developed a numerical procedure for tackling the elastohydrodynamic problem which enabled a wide range of solutions to be obtained for a study of the effect of important variables (refs. 15 and 16). This work led to an empirical relationship between minimum film thickness and the important independent variables that were found to be load, speed, atmospheric pressure viscosity at the temperature of the solids, viscosity-pressure coefficient, elastic properties of the solids, and the radius of the equivalent cylinder near a plane. This work, which provided the designer with the first useful formula for calculating minimum film thickness in gears and rolling-element bearings, was reported in references 17 and 18.

The minimum film thickness equation for elastohydrodynamic conditions predicts the existence of lubricating films 10 to 100 times greater than those estimated by conventional analysis. The lubricant film was shown to be almost constant in thickness throughout the effective load-carrying region with a restriction giving the minimum separation near the outlet end of the region. The minimum film thickness was found to be strongly dependent on the rolling speed, the viscosity at ambient pressure, and the pressure-viscosity coefficient of the lubricant, while the elasticity of the solids, the effective radius, and the load appear to play a minor role.

The theoretical hydrodynamic pressure distribution was found to be very similar to the semielliptical stress pattern of dry contact over most

of the Hertzian zone, with the exception of a sharp pressure peak that occurred toward the outlet end of the main-load supporting region in many solutions. This pressure peak is a most distinctive feature of many theoretical solutions of elastohydrodynamic problems, and it has attracted a good deal of comment. A typical elastohydrodynamic film profile and pressure distribution is shown in figure 1. It can be seen that the film has a nearly constant thickness over most of its Hertzian length, but that a restriction forming the minimum thickness occurs near the outlet end of the zone. The pressure is close to the Hertzian form apart from the sharp spike. Further theoretical solutions for isothermal conditions that demonstrated broad agreement with the work described above were produced by Archard, Gair, and Hirst (ref. 19) in 1961 and Osterle and Stephenson (ref. 20) in 1962.

The full exploitation of numerical methods in the solution of the governing equations of elastohydrodynamics that took place in the late 1950's and early 1960's was paralleled by a revolution in disc machine experiments initiated by Crook. In a remarkable series of experiments, Crook (refs. 21 to 24) made direct measurements of elastohydrodynamic oil film thickness in a disc machine by capacitance and oil flow methods. The results were in reasonable agreement with the predictions of current elastohydrodynamic theory, and the agreement dates the firm beginnings of the subject of elastohydrodynamic lubrication in the early 1960's.

Sibley and Orcutt (ref. 25) were successful in recording film thicknesses by an X-ray transmission technique in 1961, and the measurements were again in good overall agreement with theoretical predictions. The capacitance technique has been refined, and an excellent account of the experimental technique was published by Dyson, Naylor, and Wilson (ref. 26) in 1965. Optical interference methods have been employed in the measurement of elastohydrodynamic film shapes (ref. 27), and it is

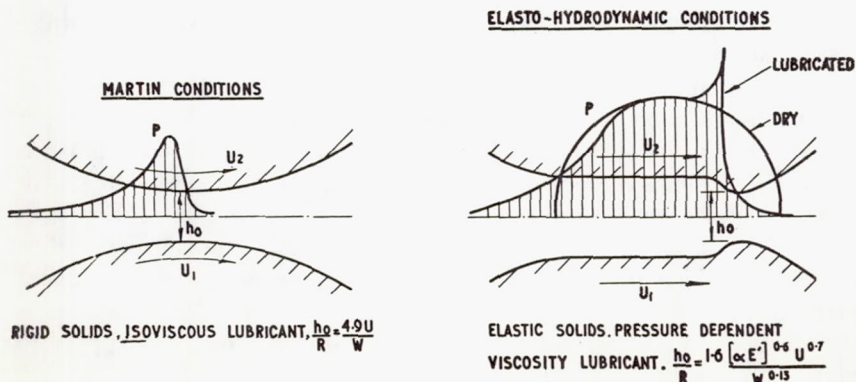


FIGURE 1.—Typical rigid and elastic cylinder pressure distributions and film shapes.



now fairly clear that the most successful experimental procedures for film thickness measurement that have been developed are capacitance, X-ray transmission, and optical interference.

Most of the initial theoretical work in the subject was concerned with isothermal and steady-state conditions for a Newtonian fluid, but some of these limitations have been relaxed in recent years. Theoretical solutions that took account of thermal effects in rolling and sliding elastohydrodynamic contacts have been presented by Cheng and Sternlicht (ref. 28) and Dowson and Whitaker (ref. 29).

The effect of non-Newtonian behavior of the lubricant has been studied by a number of research workers. One of the most satisfactory solutions has been presented by Dyson (ref. 30).

Nominal point contacts have been studied both experimentally and theoretically by Archard and Kirk (ref. 31), Archard and Cowking (ref. 32), and Cameron and Gohar (ref. 27). The apparent ease with which adequate fluid films can be established under the arduous conditions presented by point contacts is a remarkable demonstration of the importance that must be attributed to elastohydrodynamic action in many lubricated machine elements. The agreement between theory and experiment in the field of nominal point contacts has not yet reached the remarkably satisfactory state achieved for nominal line contacts; but a great deal of information has been collected, and the broad features of the phenomenon have been established.

Nonsteady-state elastohydrodynamic action has been studied for squeeze-film action by Christensen (ref. 33) and more recently by Dowson and Jones (ref. 34). It has been shown that very high localized film pressures can be generated when elastic solids approach each other in the presence of a lubricant. This action can lead to significant elastic, and maybe even plastic, indentations in the solids. The action may be important in many practical situations involving combined rolling and squeeze-film action, but no experimental or theoretical studies have been reported in this field.

The theoretical elastohydrodynamic film thickness is often very similar to the combined surface roughnesses ( $\Sigma$  rms), and the question of the effect of surface quality on effective film generation, friction, and life of lubricated contacts has recently been considered (refs. 35 to 38). The subject is sometimes called micro-elastohydrodynamics or asperity lubrication. Although the work reported to date does not provide a full picture, it does appear that local surface irregularities might produce significant effects on the overall elastohydrodynamic action.

Perhaps the greatest measure of the extent to which the subject of elastohydrodynamic lubrication has entered the broader field of tribology is provided by the application of the concept and results to practical engineering design situations. Calculation of the elastohydrodynamic



film thickness now forms an essential feature of the study of many gears, rolling-element bearings, cams, and fluid seals. In particular, it has been found that the ratio of elastohydrodynamic film thickness/ $\Sigma$  surface roughnesses has a profound effect on the life of many highly stressed lubricated contacts. Since life is often a critical aspect of machine performance, successful elastohydrodynamic behavior is being ensured at the design stage.

#### DIMENSIONAL ANALYSIS

There has been much interest in recent years in the presentation of results of elastohydrodynamic analysis. This has introduced a discussion of suitable dimensionless groups.

The major variables encountered in isothermal analysis can be represented by the following dimensionless groups:

$$\text{Film thickness parameter} \quad H = \frac{h}{R}$$

$$\text{Load parameter} \quad W = \frac{w}{E'R}$$

$$\text{Speed parameter} \quad U = \frac{\eta_0 u}{E'R}$$

$$\text{Materials parameter} \quad G = \alpha E'$$

This set of four groups can be reduced to three without any loss of generality. One convenient set is

$$\pi_1 = HG^2$$

$$\pi_2 = WG^2$$

$$\pi_3 = UG^4$$

Moes (ref. 39) has used the following formulation for representing film thickness data:  $HU^{1/2}$ ;  $WU^{1/2}$ ;  $GU^{1/4}$ .

#### FILM THICKNESS

It has already been noted that one of the most important results of recent work on elastohydrodynamic lubrication is the recognition of the importance of film thickness in determining satisfactory performance and life of lubricated concentrated contacts. It is therefore important that the operating film thickness should be capable of prediction at the design stage. Much theoretical and experimental work has been directed to this end.

The basis of most theoretical and experimental work is the observation

that many machine contacts can be represented by two cylindrical or spherical surfaces. As an example, the representation by equivalent cylinders of the contact between two involute spur gears is shown in figure 2. Once the radii of the two cylinders have been determined, the gap between the surfaces can be represented with good accuracy by the gap between an equivalent cylinder and a plane if the radius of the equivalent cylinder is determined by the relationship

$$\frac{1}{R} = \frac{1}{R_1} \pm \frac{1}{R_2}$$

The representation of contact geometry by an equivalent cylinder or sphere near a plane is convenient for the lubrication studies of highly loaded rigid components, but it is even more advantageous when elastic effects are considered. In the latter case the plane surface can be assumed to be rigid, and the cylinder or sphere can be given elastic properties that permit the sum of the elastic displacements on the original solids to be produced by the same pressure distribution on the equivalent cylinder or sphere. The cylinder or sphere near a plane is thus geometrically and elastically similar to the original contact between the machine elements. The concept can be applied with equal facility to gears, rolling-element bearings, and cams and tappets.

The object of theoretical studies in elastohydrodynamics is to obtain compatible solutions of the Reynolds and elasticity equations. If sliding action is present, the energy and heat conduction equations also have to be considered. The numerical procedures that have been found to be most fruitful are described fully in reference 1. The procedures are often

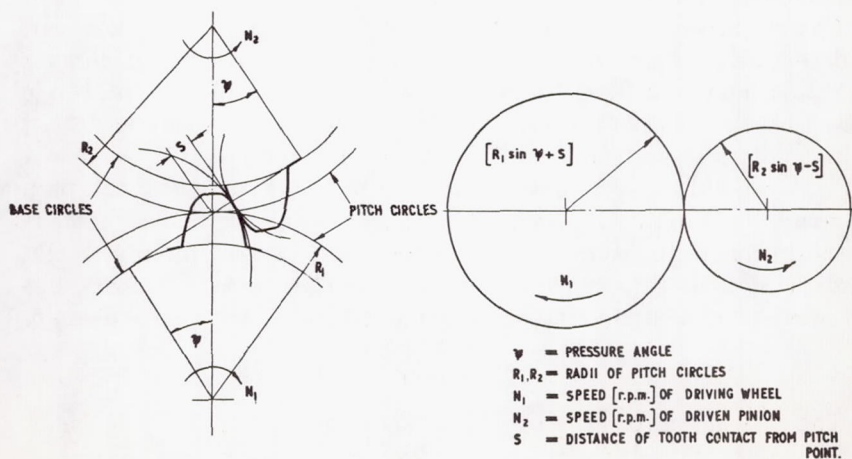


FIGURE 2.—Representation of gear tooth contact by equivalent cylinders.

difficult and time consuming, and if it were necessary to produce individual solutions for each specific contact condition, progress would be slow. However, the film thickness data are capable of representation by empirical or approximate analytical equations. This enables the designer to produce solutions applicable to his specific geometry without the need for lengthy original solutions to the governing equations.

If it is established that the contact behavior is truly elastohydrodynamic, the following film thickness equations can be used.

$$\text{Cylinder near a plane: } H_{\min} = 2.65 \frac{G^{0.54} U^{0.70}}{W^{0.13}}$$

where  $w$  = load per unit width of cylinder, and

$$W = \frac{w}{E'R}$$

$$\text{Sphere near a plane: } H = 1.40 \frac{(GU)^{0.74}}{(W')^{0.074}}$$

where  $w'$  = total load on contact, and

$$W' = \frac{w'}{E'R^2}$$

The agreement between the predictions of cylindrical and spherical contact theories and available experimental results is shown in figures 3 and 4. The experimental results were obtained by the capacitance technique on a disc machine (ref. 26) for line contacts and on a crossed cylinders machine (ref. 31) for point contacts.

Charts showing film thickness for any given values of the dominant independent variables have been constructed for classical, elastohydrodynamic, and transition or intermediate conditions. These charts are based on the relatively small number of full computer solutions from various sources. A useful chart for steel contacts lubricated by a mineral oil is shown in figure 5. A more complete representation of available data in terms of alternative dimensionless groups presented by Moes (ref. 39) is shown in figure 6. More work is required to produce the additional solutions needed for the completion of a completely satisfactory chart, but the designer can obtain reasonably sound guidance from figures 5 and 6.

#### FRICITION

The calculation of surface tractions under elastohydrodynamic conditions is more difficult than the calculation of film thickness. The reason for this is that, although the film shape can be predicted with good ac-



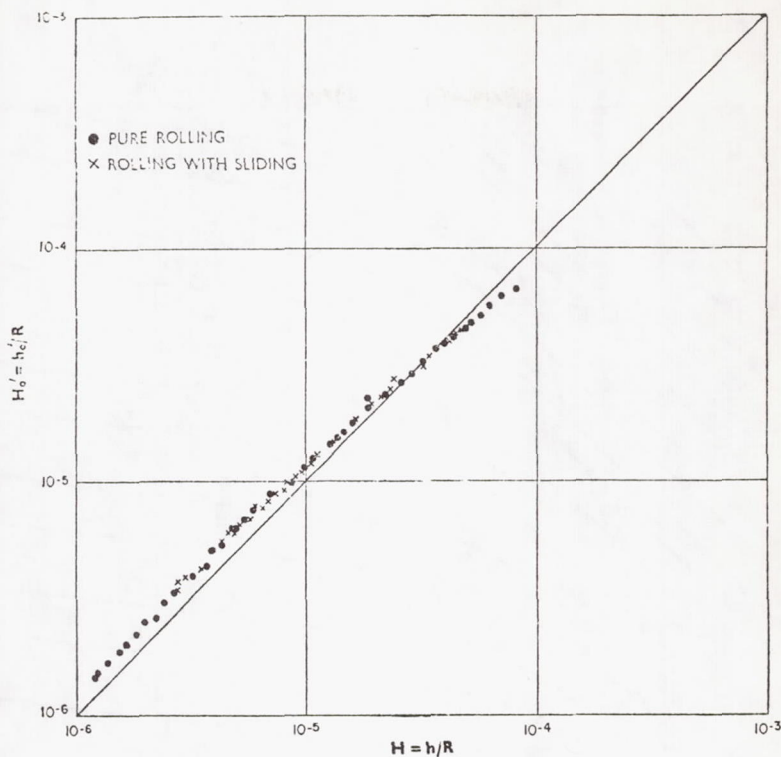
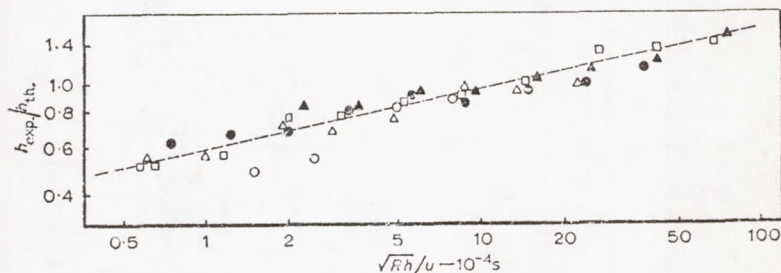


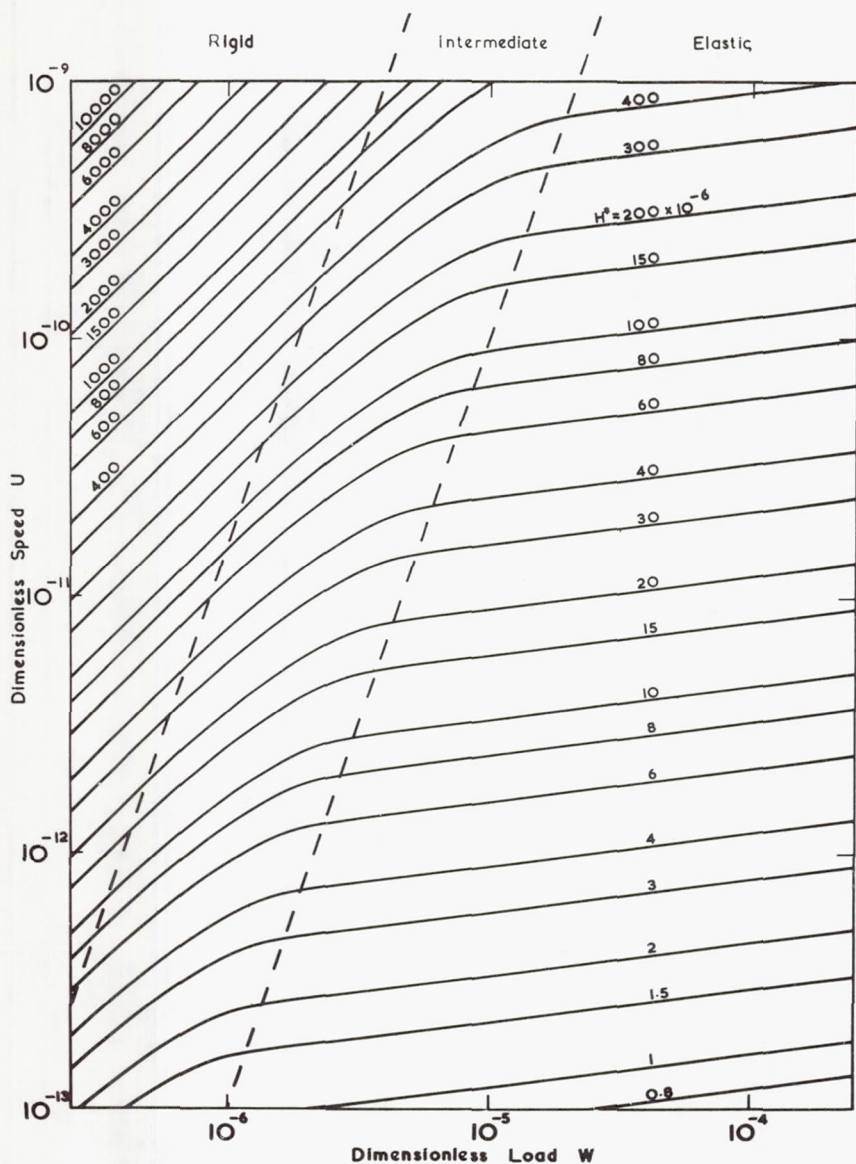
FIGURE 3.—Film thickness in line contact: comparison between theory ( $H$ ) and experiment ( $H'$ ) (after Dyson, Naylor, and Wilson, ref. 26).



Results for experiments at 20°C with crossed cylinders of various radii of curvature:  $\circ$  0.318 cm;  $\triangle$  0.635 cm;  $\square$  1.11 cm;  $\bullet$  1.97 cm;  $\blacktriangle$  3.77 cm.

FIGURE 4.—Film thickness in point contact: the discrepancy between theory and experiment (after Archard and Kirk, ref. 31).

curacy, the lubricant properties under the high-pressure, high-shear-rate, and possibly high-temperature conditions are not well defined. A typical form of friction coefficient characteristic when plotted against sliding velocity at a fixed rolling speed is shown in figure 7. If the lubricant be-

FIGURE 5.—Minimum film thickness contours ( $G = 5000$ ).

haved as a Newtonian fluid and remained at constant temperature, the relationship would be linear; but the viscosity falls due to the thermal and possibly shear-rate effects to provide the characteristic with a maximum value and then falling values of friction as the sliding velocity increases.

The viscosity of many lubricants at high pressures and a reasonable

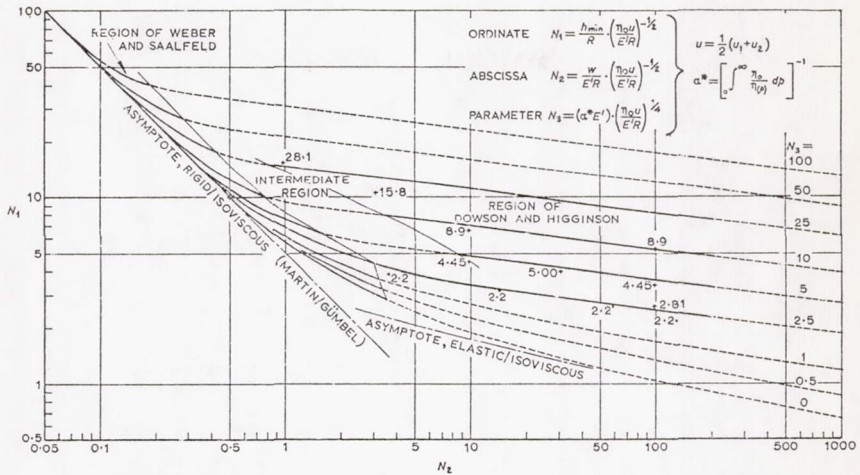


FIGURE 6.—Survey diagram for incompressible isothermal elastohydrodynamic lubrication (after Moes, ref. 39).

range of temperatures has been determined for low shear-rate conditions (ref. 40). Likewise the effect of shear rate on viscosity has been measured at atmospheric pressure (ref. 41). Until the effects of these simultaneous actions are known, it is difficult to see how satisfactory progress can be made in the understanding and prediction of friction forces under elastohydrodynamic conditions.

#### PRESSURE

The extraordinary feature of theoretical solutions to the elastohydrodynamic problem known as the pressure spike has attracted much attention in the past 10 years. A typical theoretical solution for pressure, film shape, and temperature distribution is shown in figure 8. It is a feature of undoubted scientific interest, but its significance in technological terms, particularly in relation to its influence on subsurface stresses and fatigue, has probably been overrated. The measurement of this very narrow pressure peak has presented a problem beyond the present capabilities of experimental procedures. Perhaps the most successful technique has been that developed at Battelle involving the use of a manganin strip deposited along the generator of a disc (ref. 42).

The present situation is that no measurements have recorded pressure peaks of the appropriate height and width to confirm the theoretical predictions. This position may be the result of inadequate resolution of existing experimental techniques or of an inaccurate theoretical model of the behavior of lubricants and solids, or both, and it may be some time before the position is resolved.



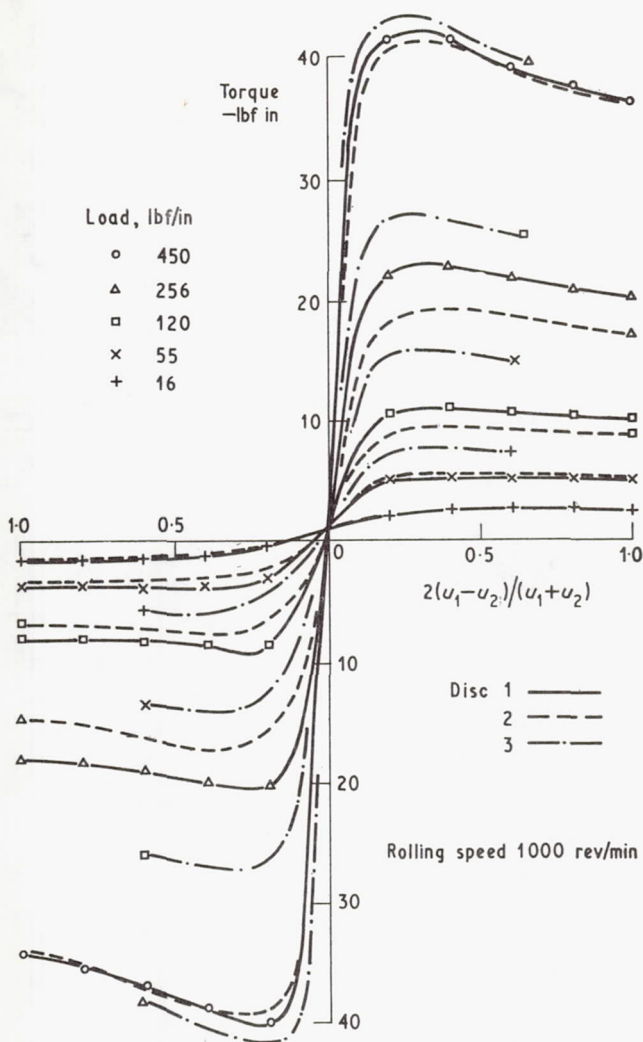


FIGURE 7.—Variation of torque with sliding speed under elastohydrodynamic conditions.

#### TRANSITION FROM HYDRODYNAMIC TO ELASTOHYDRODYNAMIC CONDITIONS

It is usually found that elastohydrodynamic effects become important in lubricated contacts when the local elastic deformation ceases to be negligible compared with the film thickness. In most cases involving metals, the effect of high pressure on lubricant viscosity becomes important before this condition is reached. In situations like soft rubber seals, however, the elastic deformation becomes large before pressure effects on lubricant viscosity need to be considered.

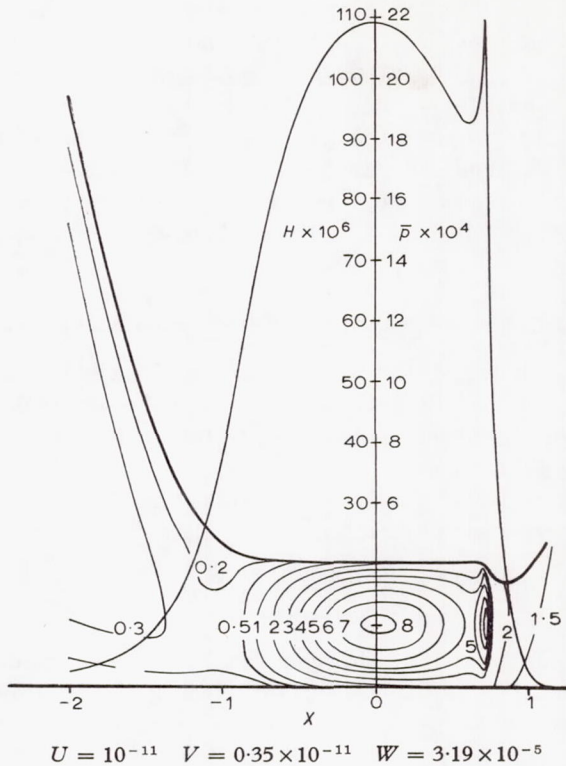


FIGURE 8.—Theoretical pressure and temperature distributions and elastohydrodynamic film shape for rolling and sliding.

#### TRANSITION FROM ELASTOHYDRODYNAMIC TO BOUNDARY LUBRICATION

This transition is usually of greater importance than the initial entry into elastohydrodynamic conditions. Many lubricated concentrated contacts operate in the "mixed lubrication" zone where the overall behavior of the contact is governed by a mixture of elastohydrodynamic and boundary lubrication effects. The subject has been discussed by Dowson (ref. 43).

It would be expected that the transition from elastohydrodynamic to boundary lubrication would be encountered when the elastohydrodynamic film thickness fell to a value equal to some measure of the surface roughness of the opposing solids and the size of the surface films of boundary lubricants.

A representative size for boundary lubricants is  $10^{-7}$  in., and surface roughnesses of machine elements presenting concentrated contacts are usually between  $10^{-6}$  and  $10^{-5}$  in. In many cases multi-molecular layers

of boundary lubricants exert an influence on contact behavior, and hence transition could occur when the effective film thickness fell to about  $10^{-6}$  to  $10^{-5}$  in.

So far we have been concerned with overall or macro-elastohydrodynamic action, but when the films become thin and transition to boundary lubrication takes place, it may be necessary to consider local or micro-elastohydrodynamic action. There is some evidence that this action is important from the work on fluid-seals. It is, therefore, a subject worthy of further study.

#### ENGINEERING APPLICATIONS OF ELASTOHYDRODYNAMIC ANALYSIS

An understanding of elastohydrodynamics emerged in the early 1960's, and we are now witnessing the application of the results of much experimental and theoretical research work to the analysis and design of machine components.

The main engineering applications so far have been concerned with

- (1) Rolling-element bearings
- (2) Gears
- (3) Fluid seals
- (4) Cams

In the rolling bearing field, complete bearing assemblies have been analyzed (ref. 44), and the important effect of elastohydrodynamic film thickness upon life of the bearing for any given surface quality has been used to produce design guides for long-life bearings. The distribution of load and minimum film thickness in a complete bearing assembly is shown in figures 9 and 10.

The elastohydrodynamic analysis of gears has been mainly restricted to pitch-point conditions (ref. 45), but some design offices now include the calculation of elastohydrodynamic film thickness in their gear design procedures. Once again the important influence of the lubricant upon life has been noted and taken into account in specifying surface quality or the lubricant. Design charts that can be used readily by the gear designer have been presented (ref. 45) in the form shown in figures 11, 12, and 13.

Most of the work on fluid seals and cams comes within the field of analysis, but there is every prospect that elastohydrodynamics will find a place in future design procedures.

#### FILM THICKNESS/ROUGHNESS RATIO

The ratio of effective elastohydrodynamic film thickness/surface roughness has been shown to have a very important influence on the life of lubricated concentrated contacts (refs. 46 to 48). This is probably one of the most significant results of elastohydrodynamic analysis. It affords an opportunity for the designer to optimise life of lubricated concentrated



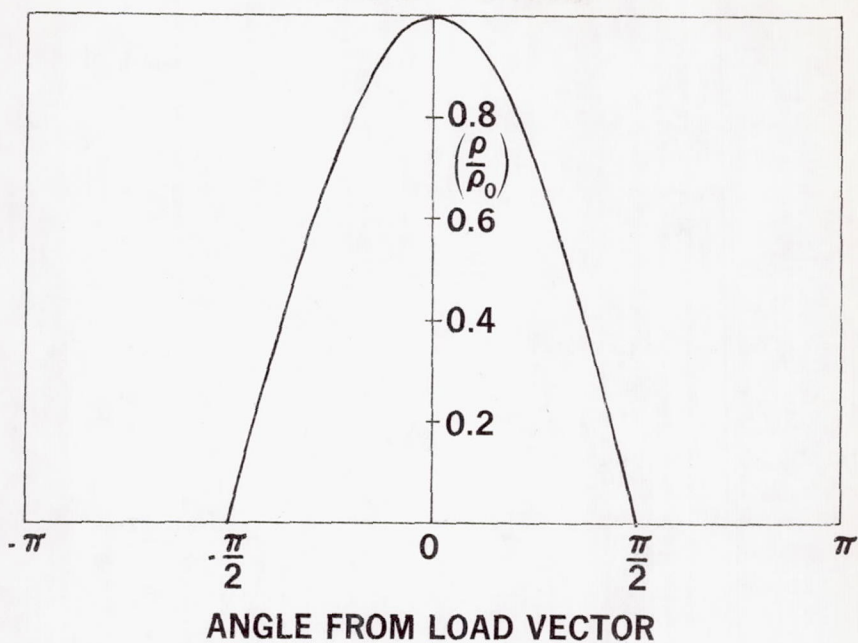


FIGURE 9.—Load distribution in a roller bearing.

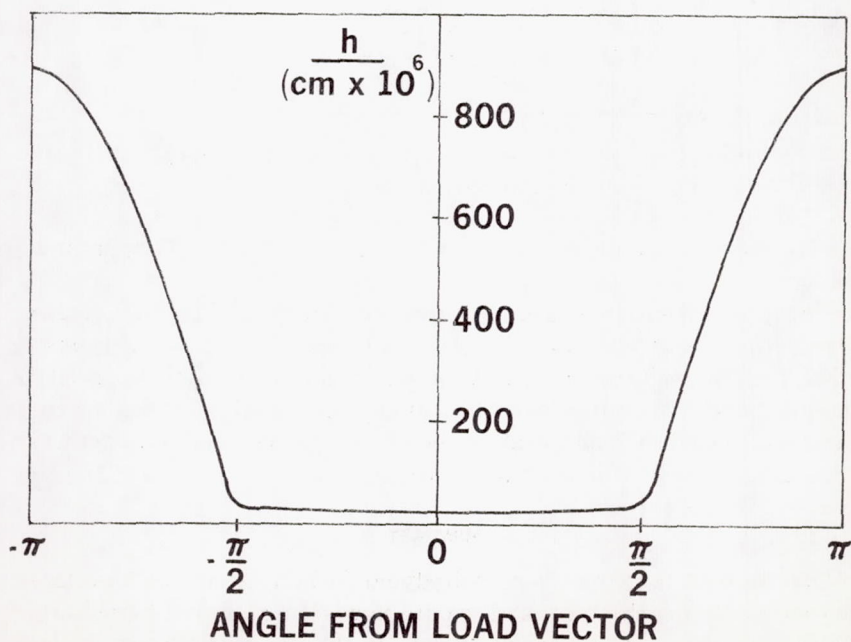


FIGURE 10.—Variation of theoretical film thickness in a roller bearing.

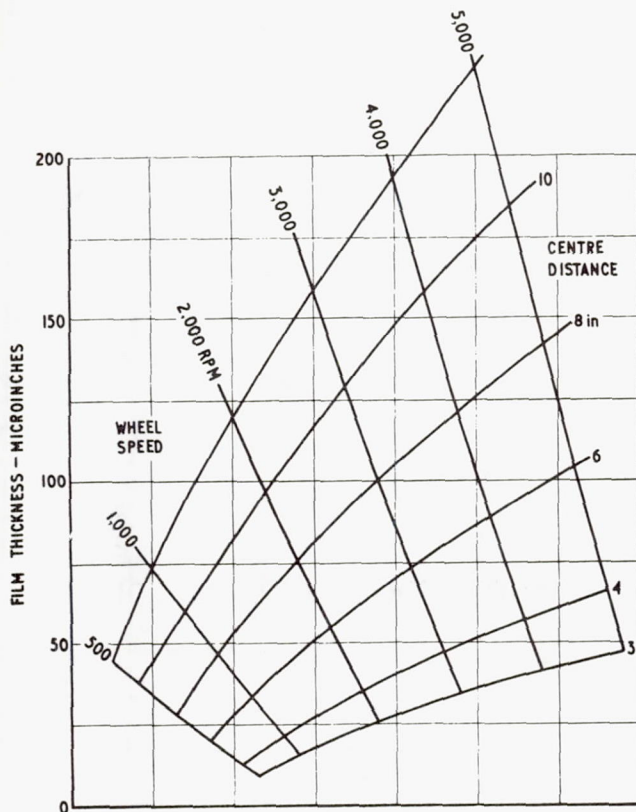


FIGURE 11.—Pitch-line film thickness for gear ratio = 1, load = 1 ton/inch, viscosity = 0.75 poise.

contacts by providing a suitable lubricant or surfaces of appropriate quality.

The film thickness is usually calculated from available expressions (see earlier), and the sum of the roughnesses is presented in terms of c.l.a., rms, or peak-to-valley. It does not seem to be possible to specify a unique value of the ratio that will optimise life for all machine elements, but ratios between 2 and 5 have been found to be suitable in terms of c.l.a. measurements for most applications.

#### SUMMARY

The present work on elastohydrodynamic lubrication has developed beyond the research stage, and we are now seeing the results of earlier studies incorporated in machine design. In the coming years we shall see further and more complete elastohydrodynamic studies of lubricated concentrated contacts which will provide a firmer basis for design.

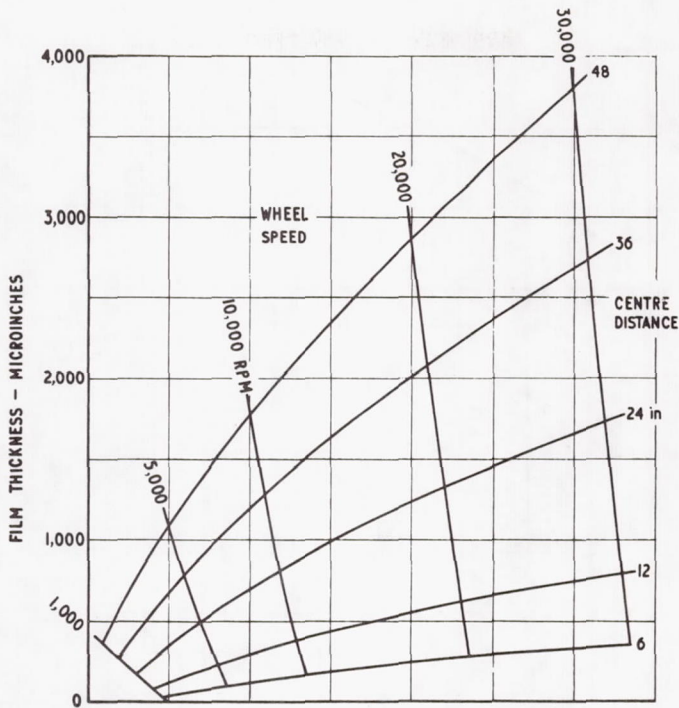


FIGURE 12.—Pitch-line film thickness for gear ratio = 1, load = 1 ton/inch, viscosity = 0.75 poise.

The main area requiring work is the field of mixed lubrication where elastohydrodynamic and boundary lubrication effects are both apparent. The study of elastohydrodynamics is still limited by a lack of understanding of lubricant properties under conditions of high pressure and shear rate at various temperatures. It is also becoming urgent for more realistic representation of the surface motions in machine contacts to be included in analysis. In particular, squeeze film action may play a significant role in gear and rolling bearing lubrication. There is little information on the effects of starved lubrication conditions that are also encountered in some elastohydrodynamic situations.

A new application which may be the subject of activity in the near future is in the field of metal forming. Most analyses of lubrication in metal-forming processes consider a rigid/plastic model for the worked material, but this may be inadequate in the inlet and exit regions. This new subject has been termed elasto-plasto-hydrodynamic lubrication.\*

\* Bloor, S.; Dowson, D.; and Parsons, B. P.: J. Mech. Eng. Sci., Vol. 12, 1970, p. 178.



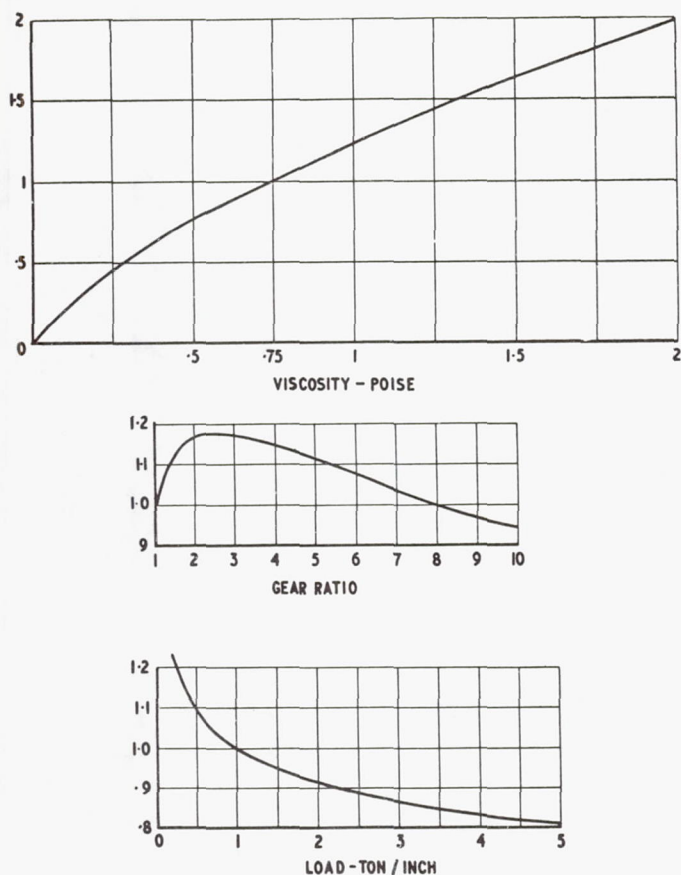


FIGURE 13.—Correction factors for centerline film thickness.

There is still a great deal to be done in connection with fundamental studies and the application of elastohydrodynamic principles, but it is satisfying to note the extent to which earlier research work is now influencing the design of some machine components. It appears that the generation of an effective elastohydrodynamic film is a necessary, but perhaps not complete, requirement for satisfactory life and performance of many lubricated concentrated contacts.

#### ACKNOWLEDGMENTS

Figures 1, 2, 11, 12, and 13 have been reproduced from reference 45 by kind permission of the Institute of Petroleum; figures 3, 4, 6, 7, and 8 are reproduced by kind permission of the Institution of Mechanical Engineers; figure 5 is reproduced by kind permission of the American

Society of Lubrication Engineers; and figures 9 and 10 were published in reference 44.

#### DISCUSSIONS

##### A. Dyson (Shell Research Limited, Thornton Research Center, Chester, England)

This symposium was to have included an invited lecture on Elastohydrodynamic Lubrication by Professor D. Dowson. Unfortunately Professor Dowson was ill and unable to attend the conference, and the writer was asked to read Professor Dowson's paper to the symposium. The preprint submitted by Professor Dowson was not intended for verbal presentation, and did not include any illustrations. It was therefore necessary for the present author to give his own interpretation of Professor Dowson's subject matter and to assemble his own illustrations. The result inevitably differed to some extent from the original, and the author has now been asked to submit a written version for inclusion in the proceedings of the symposium.

An apology must be made for the fact that the majority of the illustrations derive from the author's country and that many of them are from his own work. The reason for this is that the selection of illustrations had to be made on rather short notice, and it was inevitable that those most readily accessible were chosen.

After a brief historical review, the present state of the art of elastohydrodynamics is discussed, and examples are given of some recent advances in theoretical analysis and in experimental technique. Some examples of the application of elastohydrodynamic theory to practical questions are given, and finally the problems to be solved in the future are enumerated.

*History*—The history of the subject has been covered adequately by Professor Dowson in his paper, and only a brief review will be given here. One of the important characteristics of an elastohydrodynamic system is that in general the surfaces are counterformal; that is, their centers of curvature lie on opposite sides of their conjunction. Most such systems of practical interest may therefore be simulated by the external contact between two rollers or discs. This model has been used in both theoretical and experimental work.

The first important theoretical analysis of this situation was made by Martin (ref. 49) who considered the classical hydrodynamic case of rigid surfaces separated by an isoviscous lubricant. The minimum thickness of the hydrodynamic film between gear teeth, calculated from Martin's theory, was of the order of 25 nm (1 microinch), which is only about one-hundredth of the peak-to-valley roughness of the surfaces commonly employed. Such surfaces still displayed their original grinding marks after thousands of hours of operation, and the Martin theory was obviously inapplicable.

Attempts were made to extend the theory by the separate inclusion of two effects of the high pressures necessarily developed in the conjunction—the increase of the viscosity of the lubricant and the elastic deflections of the surfaces. It was not until both effects were considered simultaneously that calculated film thicknesses consistent with the results of experience were obtained. An example of the interaction between the elastic deflections and the effect of pressure on viscosity is given in figure 14, reproduced from a paper by Dowson and Higginson (ref. 14). To obtain the results shown in figure 14, the centerline film thickness, the viscosity of the lubricant at atmospheric pressure, and the rolling speed were assumed to be constant. The load-carrying capacity, i.e., the area under the pressure distribution curve, is increased only slightly if the effect of pressure on the viscosity or the elastic deflections are considered singly, but the effect is much greater when both these factors are allowed to operate simultaneously.

An important development in the theoretical understanding of the

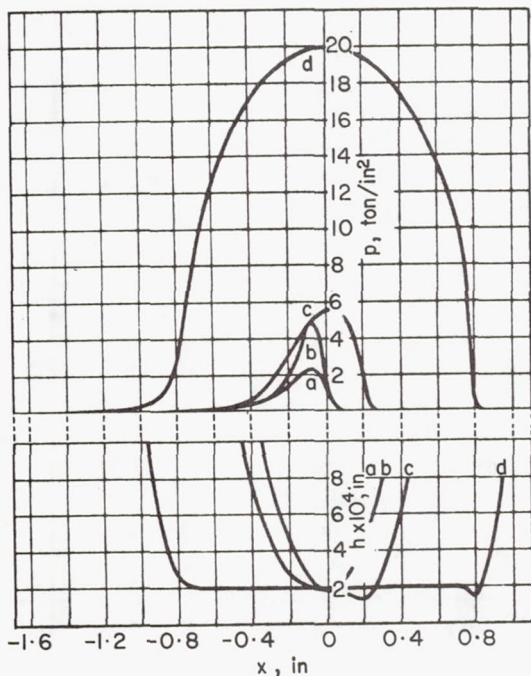


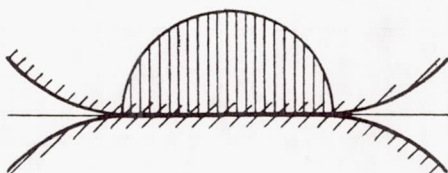
FIGURE 14.—Pressure distributions and film shapes for the same centerline film thickness: (a) constant viscosity, rigid cylinders; (b) pressure-dependent viscosity, rigid cylinders; (c) constant viscosity, elastic cylinders; and (d) pressure-dependent viscosity, elastic cylinders.



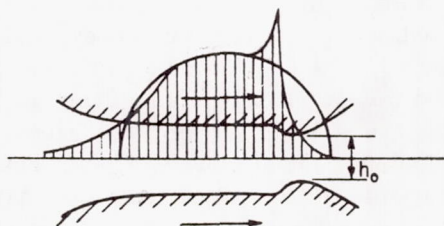
subject was introduced by Grubin (ref. 50), who used a powerful combination of physical insight and simple mathematical arguments to reveal many of the essential features of elastohydrodynamic lubrication. In particular, he described the film shape, including the essentially parallel central portion and the outlet hump, evident in figures 15 and 16. He also produced a general formula that gives a very good approximation to the central film thickness. Both these results have been confirmed by the more direct but laborious procedure of obtaining simultaneous solutions to the hydrodynamic and elastic equations governing the behavior of the system. Petrusevitch (ref. 51) obtained such solutions, but only for three special cases. A more generally applicable solution was given by Dowson and co-workers (refs. 14 and 18). A distinguishing feature of the work of the latter authors is their use of inverse hydrodynamic theory to



a. Martin conditions — rigid solids, isoviscous lubricant



b. Hertzian conditions — dry contact, elastic solids



c. Elastohydrodynamic conditions  
—elastic solids, Newtonian lubricant

FIGURE 15.—Contact conditions.

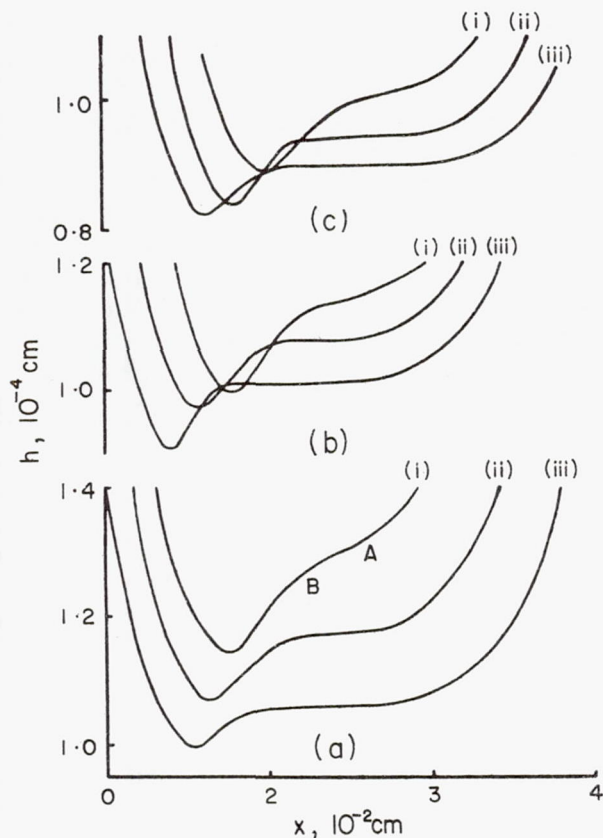


FIGURE 16.—Film shapes [rolling speeds: (a) 25; (b) 9.2; (c) 3.3 ft s<sup>-1</sup> and loads: (i) 230; (ii) 360; (iii) 500 lb in<sup>-1</sup>].

obtain a convergent solution. The essential features of these solutions are illustrated in figure 15 (ref. 3). The excellent general agreement of these theoretical predictions with the results of experiments is illustrated in figure 16 reproduced from a review given by Crook (ref. 52), and in figure 17 reproduced from a paper by Kannel (ref. 42).

Dowson, Higginson, and Whitaker (ref. 18) gave a useful empirical equation for the minimum film thickness which summarizes the results of a large number of numerical solutions that they have obtained:

$$\frac{h_{\min}}{R} = 1.6 \left( \frac{\eta_0 \bar{U}}{E'R} \right)^{0.7} (\alpha E')^{0.6} \left( \frac{w}{E'R} \right)^{-0.13} \quad (1)$$

where  $h_{\min}$  is the minimum thickness,

$R$  is the radius of relative curvature, given by

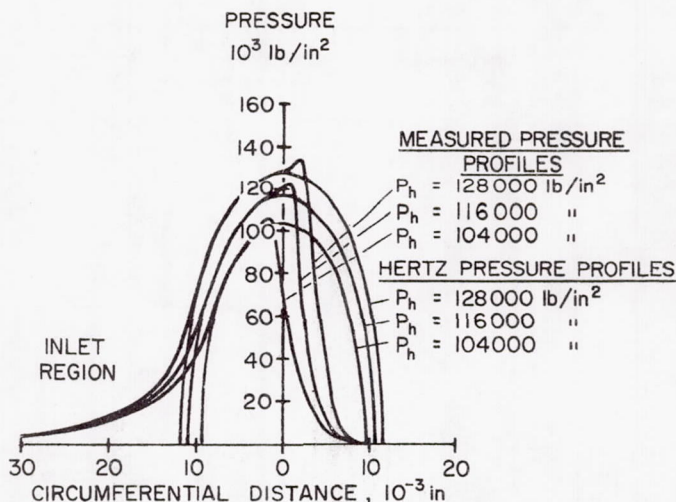


FIGURE 17.—Measured variation in pressure profile with disc-loading.  
 $T = 115^\circ \text{ F}$ ,  $RS = 1400 \text{ ft/min}$ , lubricant—polyphenyl ether.

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

$R_1$  and  $R_2$  are the radii of curvature of the two surfaces,  
 $\eta_0$  is the viscosity of the lubricant at atmospheric pressure and at  
 the temperature of the surfaces as they enter the conjunction,  
 and

$\bar{U}$  is the rolling speed.

The results were derived for pure rolling. For moderate degrees of sliding, the mean of the peripheral velocities of the two surfaces, relative to their conjunction, may be used.

$E'$  is the reduced elastic modulus, given by

$$\frac{2}{E'} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2}$$

where  $\nu_1$ ,  $\nu_2$  and  $E_1$ ,  $E_2$  are the Poisson's ratios and the Young's moduli of the materials of the two surfaces, 1 and 2;

$\alpha$  is the pressure coefficient of viscosity, defined by the relation

$$\eta(P) = \eta_0 \exp(\alpha P)$$

$\eta(P)$  is the viscosity at pressure  $P$  and at the temperature of the surfaces entering the conjunction; and

$w$  is the load per unit width of contact, measured in a direction transverse to that of rolling.



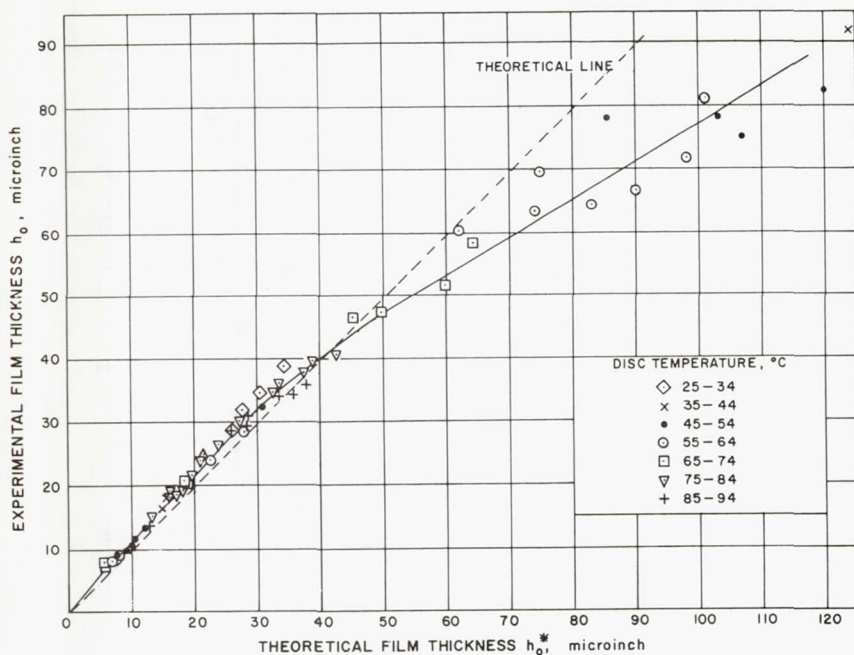


FIGURE 18.—Comparison of experimental film thickness, as deduced from capacity measurements, with theory, oil A, sliding, velocity ratio 33/63.

The agreement between equation (1) and the experiment is illustrated in figure 18 (ref. 53). The experimental film thicknesses were estimated from measurements of the electrical capacitance between two rolling and sliding discs loaded together. The agreement is excellent for thin films up to about 1 nm (40 microinches) thick, although the thinnest films are slightly thicker than predicted by the theory. This probably arises because the theory gives the minimum thickness under the outlet "bulge," while the experiments give a harmonic mean over the conjunction region. On the other hand, the thicker films are thinner than the theory predicts, and the reason probably lies in the heating by shear of the lubricant in the inlet film. This factor is ignored in the derivation of equation (1), and theories in which this viscous heating is taken into account in an approximate manner show a similar falling off in film thickness from the isothermal value.

Some further results\* for very thin films are shown in figure 19, and two points of interest emerge. The first is that the thinnest film measured was about 10 nm (100Å), or only a few molecular dimensions thick, but

\* Data by A. R. Wilson, unpublished.

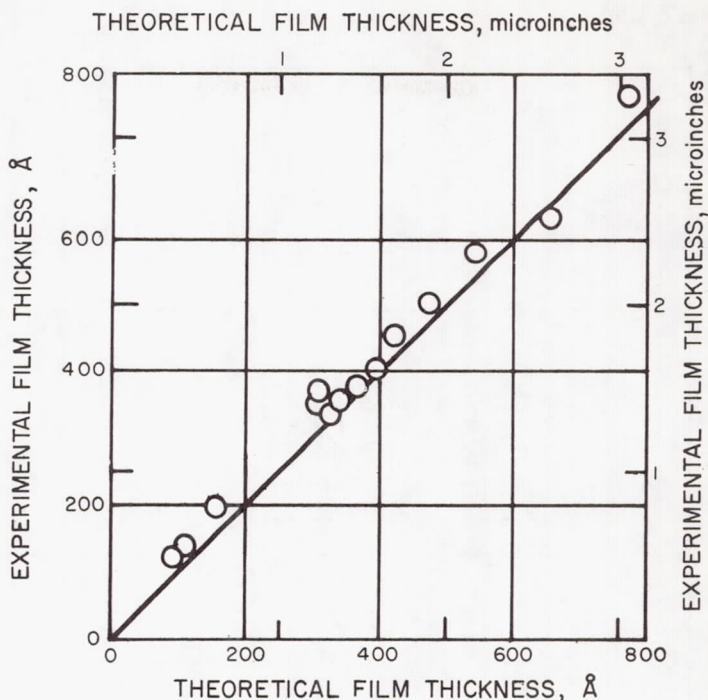


FIGURE 19.—Comparison of theoretical and experimental film thicknesses for very thin films.

that nevertheless the thickness was predicted well by theories based on continuum mechanics. The second point is that under these conditions, there is no trace of the long-range influences that many authors have postulated and that are claimed to enhance the viscosity of a fluid near a metal surface. Theoretical predictions based on the viscosity of the fluid in bulk agree well with the results of experiment.

*Present state of the art*—The quantities of most interest in any lubricated system are the minimum film thickness, the frictional traction, and the design limits for satisfactory operation. Each of these will be discussed, but most of the emphasis will be on the minimum film thickness. The other two topics will receive only brief mention.

The good agreement between the theoretical equation (1) and experimental results for the minimum film thickness, and the discrepancy in the thick-film region, have already been pointed out. Further discrepancies occur at very high slide/roll ratios. These are illustrated in figure 20 (ref. 54). The slide/roll ratio is defined as

$$\frac{U_1 - U_2}{\bar{U}}$$

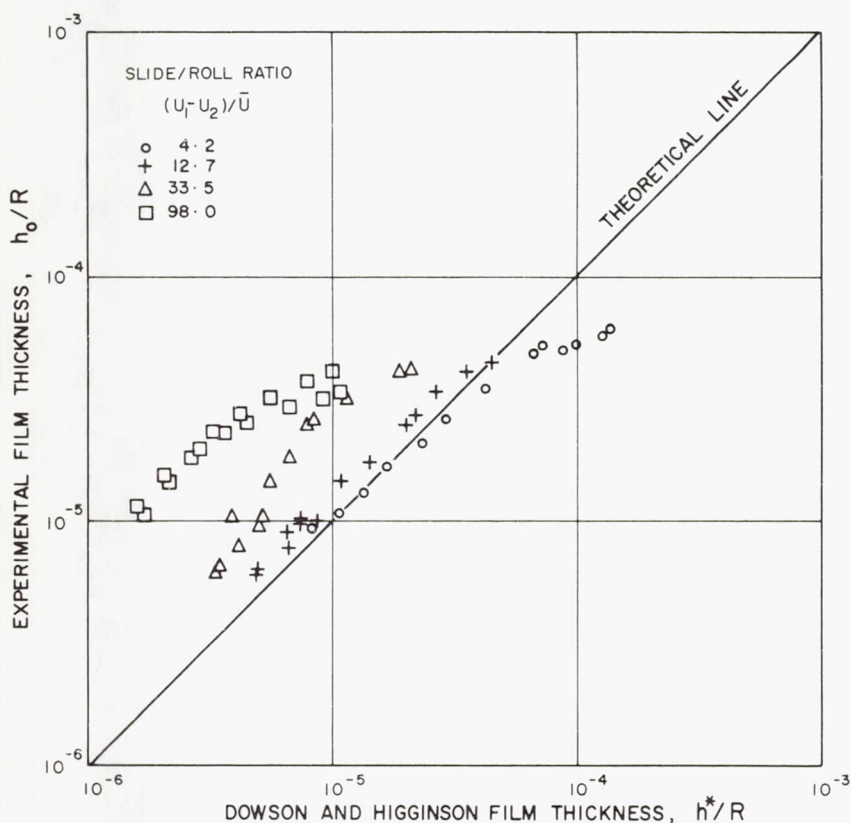


FIGURE 20.—Comparison of measured non-dimensional oil film thickness,  $h_0/R$ , with predicted values,  $h^*/R$ , at high slide/roll ratios: load, 2.45 kN (550 lbf).

where  $U_1 - U_2$  is the sliding speed and  $\bar{U} = \frac{1}{2}(U_1 + U_2)$  is the mean rolling speed,  $U_1$  and  $U_2$  being the peripheral velocities of the surfaces of the discs relative to their conjunction. A slide/roll ratio of 2 is obtained with one disc stationary. Values greater than 2 imply that the peripheral velocities of the discs are directed in opposite senses. The discs are often described as "chain-connected" in this condition. The measured film thicknesses are up to an order of magnitude larger than the predictions of equation (1). In the extreme case the discs are operated with equal and opposite peripheral velocities. Since  $\bar{U} = 0$ , the conventional theory implies the complete absence of a hydrodynamic film. The measured values of film thickness were quite large (several tenths of  $1 \mu\text{m}$ ), and quite heavy loads were carried (up to 1600 lb on a contact  $\frac{7}{8}$  in. wide between two steel discs each of 3 in. diameter). Coefficients of friction ranged down to 0.015, and the lubrication was obviously hydrodynamic.



The explanation starts with the fact that the metal surface emerging from the conjunction is hotter than that entering it. There is a difference in the properties, and particularly the viscosity, of the lubricant across the film. This difference introduces a new term in the Reynolds' equation whose effect can be quite powerful in the absence of the conventional geometric-wedge action. The principle of this action was explained by Cameron (ref. 55), but he applied it to a classical theory of hydrodynamic lubrication. A complete solution for the elastohydrodynamic case has not been obtained, but two extreme solutions have been identified, between which any practical case must lie. Some results are illustrated in figure 21 (ref. 54) in which the experimental results are seen to be within the extremes predicted by the theory. The modified Dowson and Higginson film thickness is the thickness which would be obtained according to equation (1) if the velocity of one of the surfaces

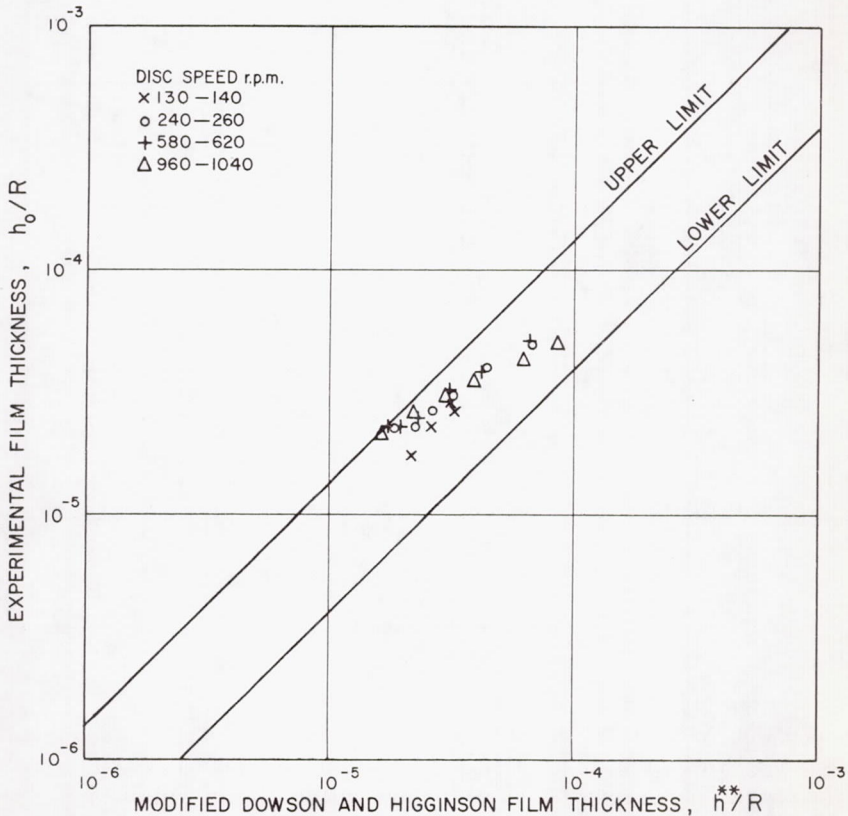


FIGURE 21.—Comparison of measured oil film thicknesses with a modified Dowson and Higginson expression: load, 7.12 kN (1600 lbf);  $U_1 + U_2 = 0$ .

were reversed, i.e., if the discs were to be operated under the conditions of pure rolling instead of under those of pure sliding with no rolling component. It is a surprising feature of the theory, confirmed by experiment, that films formed with equal and opposite peripheral velocities may in some circumstances be at least as thick as those formed in pure rolling under otherwise identical conditions. Another surprising result of the theory is that under certain circumstances the film will become thicker with increasing load. This is confirmed by the experimental results shown in figure 22 (ref. 54).

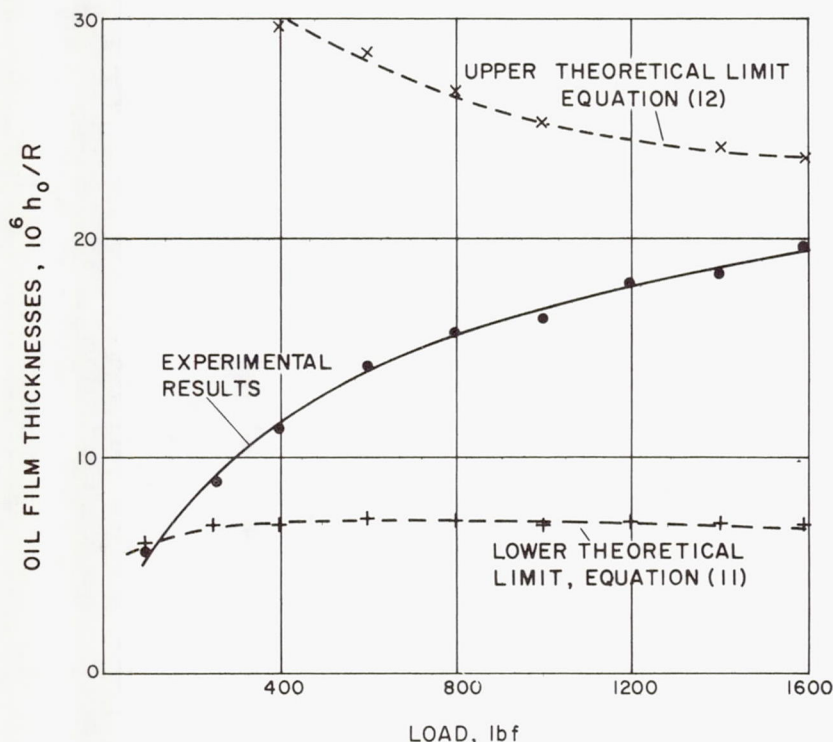


FIGURE 22.—Effect of load on measured oil film thickness (indicated disc temperature, 80° C; sliding speed  $U_1 - U_2$ , 480 cm/sec).

Further deviations from equation (1) are observed when the lubricant is not a Newtonian fluid. Dr. Naylor treats this question of lubricant rheology in his paper; therefore, the present discussion will be limited to the consideration of two examples. The first concerns the silicones, poly(dimethylsiloxanes). These are known to exhibit non-Newtonian behavior—their viscosities start to decrease with increasing shear rate

at shear rates ( $10^2$  to  $10^4$  sec $^{-1}$ ) several orders of magnitude below the maximum values obtained in disc machines ( $10^6$  to  $10^7$  sec $^{-1}$ ). It is very difficult to obtain direct information about the viscometric properties of liquids at such high shear rates in conventional viscometers. The theory may be extended to cover this case by the following procedure. The non-Newtonian behavior of silicones is attributed to their viscoelasticity. Information about their viscoelastic properties is obtained from measurements in oscillatory shear when the liquid is in contact with a piezoelectric crystal oscillating in shear (ref. 56). The behavior of the liquids in continuous shear at high shear rates may be predicted approximately from their behavior in oscillatory shear at high frequencies. It turns out that the poly(dimethylsiloxanes) behave approximately as power-law fluids with a shear stress proportional to the shear rate raised to some power not equal to unity; values of about 0.6 have been derived from the experimental results. A theory for the film thickness in elastohydrodynamic lubrication of the Grubin type may then be worked out, with this power-law replacing the Newtonian expression for the shear stress.

Figure 23 (ref. 57) shows a comparison between the experimental results and various alternative formulations of the theory based on the viscoelastic analogy. The theory gives the correct order of magnitude,

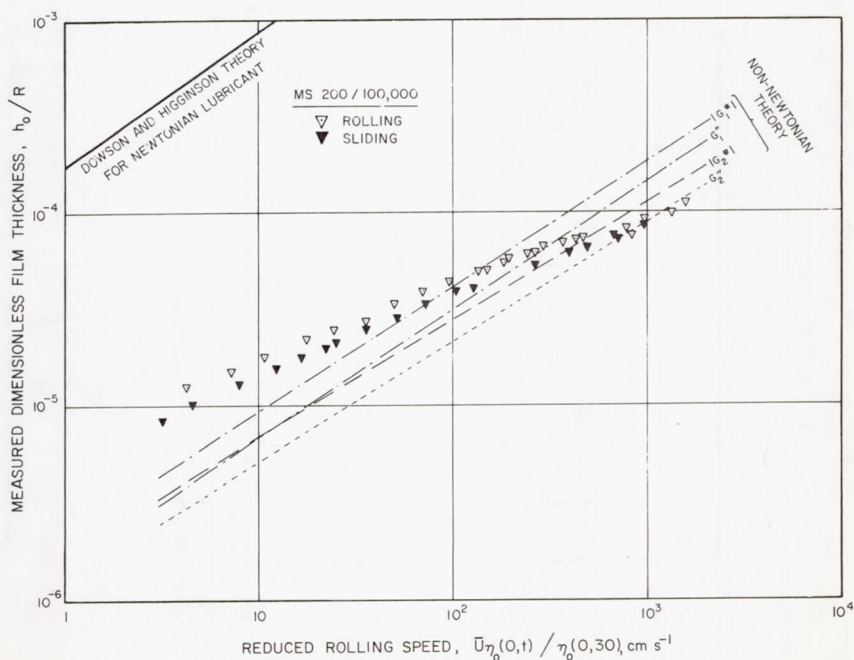


FIGURE 23.—Comparison of observed and predicted film thickness, 100 000 cs silicone.



and the error in slope is believed to be due to the inadequacy of the power-law relation over the whole range of shear rates required. The theory for a Newtonian lubricant, based on the viscosity measured in a capillary viscometer at low shear rates, gives results too large by nearly two orders of magnitude.

In the absence of thermal effects, the thickness of the film generated by a Newtonian lubricant should depend only on the mean rolling speed

$$\bar{U} = \frac{1}{2}(U_1 + U_2)$$

and should be independent of the sliding speed,  $(U_1 - U_2)$ . This result is not true for a non-Newtonian fluid, and an approximate theory for the effect of sliding speed on film thickness may be worked out for a power-law fluid. The results are compared with the experiment in figure 24 (ref. 57), and the agreement is seen to be excellent. It would naturally be expected that the sharp corner in the theoretical curve at a slide/roll of 0.83 would be rounded off in a more adequate theory.

The second class of non-Newtonian lubricants in common use is that of the greases. It seems likely that the three-dimensional fibrous structure associated with the gelling agent will be broken up at the high shear rates obtained in elastohydrodynamic lubrication and that the effective lubri-

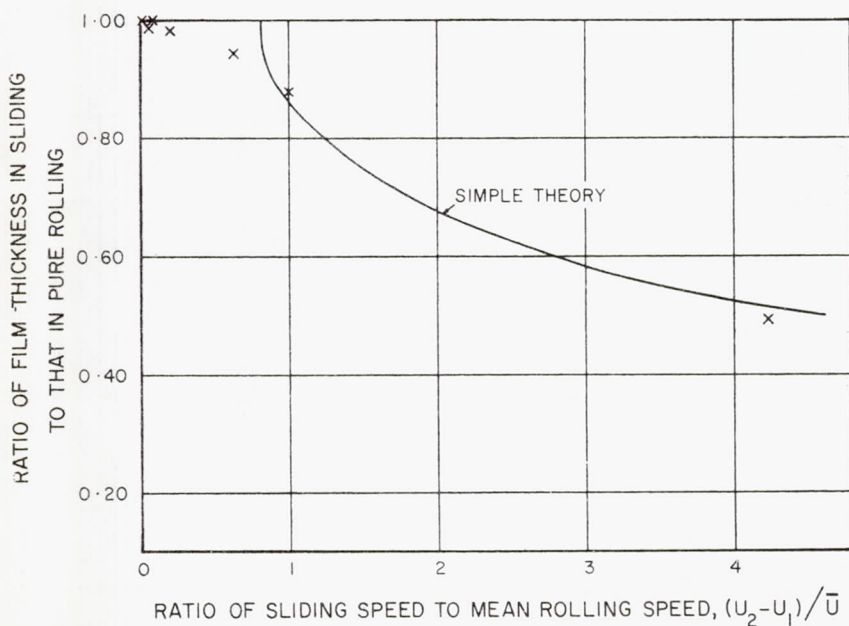


FIGURE 24.—Effect of sliding on film thickness at constant rolling speed (1000 cs silicone,  $\bar{U} = 50$  cm/s, disc surface temp.  $30^\circ$  C).

cant will be the base oil in which the gelling agent is incorporated. It therefore seems appropriate to predict film thicknesses in grease-lubricated bearings by using the properties of the base oil in the appropriate formulae, e.g., equation (1).

Some experiments we have carried out show that this procedure may lead to serious over-estimates of the film thickness (ref. 58). We put a lump of grease (approximately 0.5 ml in volume) into the inlet zone of a contact between loaded rollers and measured the film thickness as a function of time. We compared the results with those obtained by feeding a single drop of oil (approximately 0.05 ml in volume) into the gap from the end of a glass rod. It is emphasized that the lubricants were applied once only and were not replenished during the tests.

Some typical results are shown in figure 25 (ref. 58). The oil alone gave a film whose thickness did not vary with time. The grease film was initially thicker than the oil film but rapidly became thinner. By the end of the run, the grease films were only about half as thick as the oil films. It is suggested that these results may be explained by the starvation of the inlet section as a consequence of the normal stress differences and of the stored elastic energy present in a sheared viscoelastic medium such as a grease. Some support for this explanation was obtained from the behavior of a solution of polyisobutene in a mineral oil. This solution would also have viscoelastic properties. It behaved like a grease, in that it formed a film which was initially thicker than that given by the base oil

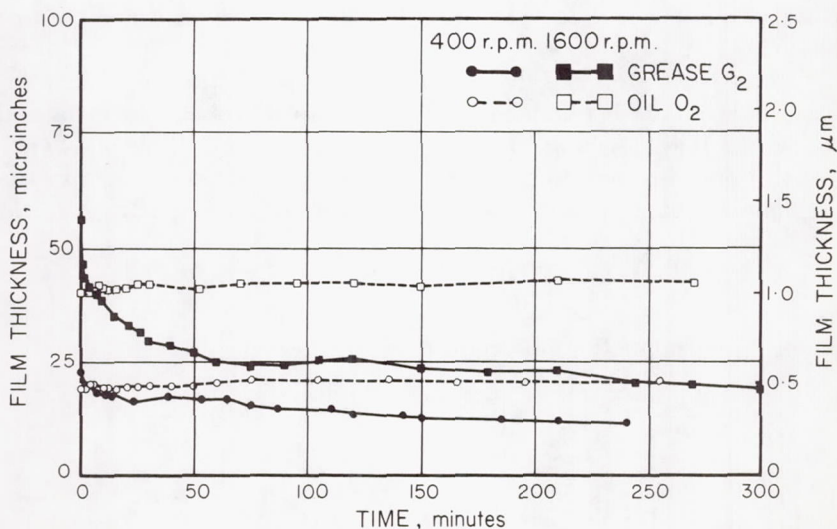


FIGURE 25.—Variation with time of film thickness of grease  $G_2$  and base oil  $O_2$  (60° C, 2200 lbf load).

above, but which rapidly became thinner than that formed by the oil. These results are shown in figure 26.

Further differences between greases and oils are illustrated in figure 27. In these experiments the discs were run at speeds of 1600 and 400 rpm during alternate periods of 15 minutes. The thickness of the film formed by the oil always changed with the speed changes, but the effect of the speed changes on the thickness of the grease film gradually decreased. Finally the thickness of the grease film became constant, independent of speed, and approximately equal to that which would have been obtained if the discs had been run continuously at the lower speed.

This discussion of the present state of the elastohydrodynamic art has concentrated on film thickness since this is the subject on which most progress has been made. Of the other quantities of importance, frictional traction and failure limits, there is little which can be usefully said in this paper. Many authors have published experimental measurements of frictional traction, and in general there are as many empirical correlations for friction as there are papers published on the subject. There is certainly no established theory of a status comparable with that of the film thickness theory. It seems in this field that the assumption of a Newtonian lubricant must be abandoned, even for straight mineral and synthetic oils, and even when thermal effects are included in the theory. In a contribution to Dr. Naylor's paper, I intend to advance a theory based on the viscoelastic properties of mineral oils in an attempt to explain some of the friction results obtained in certain conditions. The subject of failure limits is treated by many other authors at this symposium, and it would be superfluous to enter into any discussion of this phase in the present paper.

In this paper it is possible to mention only selections of recent advances that have been made in theoretical analysis and experimental technique. It has been pointed out (ref. 59) that only three nondimensional groups are necessary to characterize an isothermal elastohydrodynamic situation, instead of the four used by Dowson and Higginson:

$$\frac{h_{\min}}{R} \left( \frac{\eta_0 \bar{U}}{E'R} \right)^{-1/2}; \quad \left( \frac{w}{E'R} \right) (\eta_0 \bar{U}/E'R)^{-1/2}; \quad (\alpha E') \left( \frac{\eta_0 \bar{U}}{E'R} \right)^{1/4}$$

or suitable combinations of products of these groups (ref. 60), e.g.,

$$\left( \frac{h_{\min}}{R} \right) (\alpha E')^2; \quad \left( \frac{\eta_0 \bar{U}}{R} \right) (\alpha E')^3; \quad (\alpha E'^2) \left( \frac{w}{E'R} \right)$$

$$\left( \frac{h_{\min}}{R} \right) \left( \frac{\eta_0 \bar{U}}{w} \right)^{-1}; \quad (\alpha E') \left( \frac{w}{E'R} \right)^{1/2}; \quad (\alpha E') \left( \frac{w}{E'R} \right)^{3/2} \left( \frac{\eta_0 \bar{U}}{E'R} \right)^{-1/2}$$

Another promising advance was made by Herrebrugh (ref. 61), who



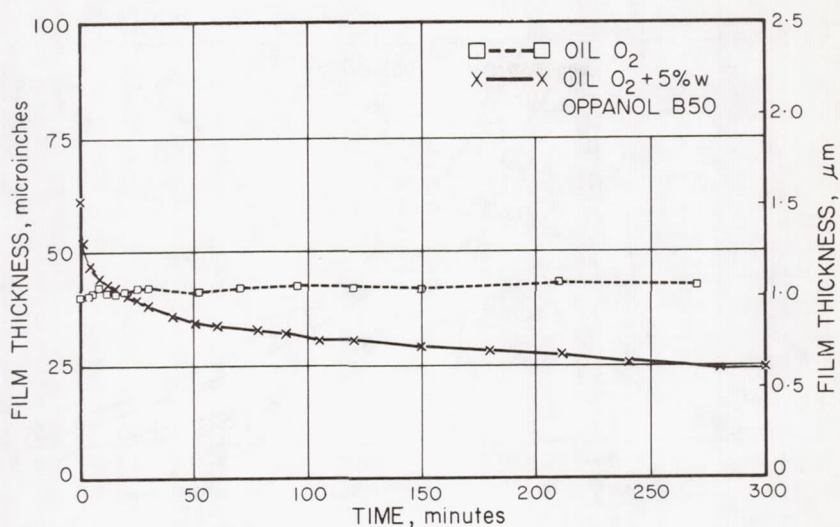


FIGURE 26.—Effect on film thickness of the addition to the test oil of 5% Oppanol B50 (60° C, 2200 lbf load, 1600 rpm).

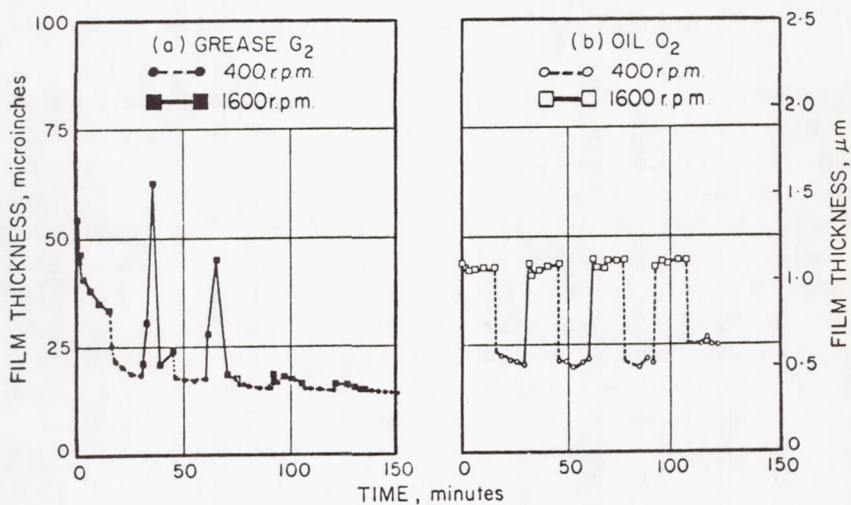


FIGURE 27.—Effect of variation of speed on film thickness of (a) grease  $G_2$  and (b) base oil  $O_2$  (60° C, 2200 lbf load).

coupled the elastic and hydrodynamic equations governing the behavior of an elastohydrodynamic system, forming a single, difficult integral equation.

On the experimental side, a new technique of obtaining the pressure

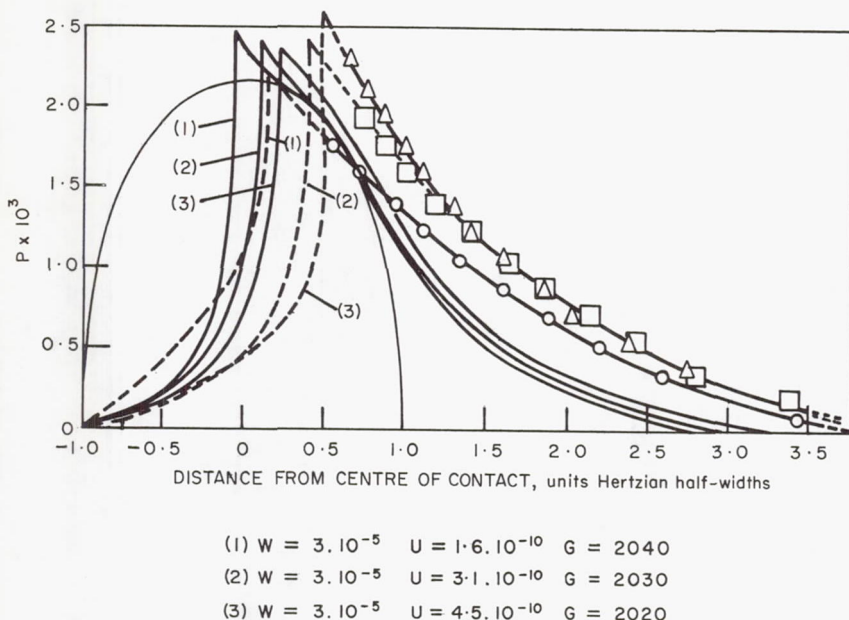


FIGURE 28.—Comparison of theoretical and experimental pressure distributions.

distribution by an optical method has been developed.\* The usual photoelastic technique permits the evaluation only of the principal stress differences, but Cope and Haines modified the technique to obtain sufficient information to establish the individual principal stresses and, therefore, the compressive stress in lubricated glass rollers. They did this by forming interference patterns with four light beams obtained from a laser source, the rays polarized in two directions at right angles and reflected from two surfaces of the specimen. In this way the intensity distribution of the isochromatic pattern obtained in conventional photoelasticity, describing contours of constant principal stress difference, is modulated by another pattern describing contours of constant principal stress sum. An example of their results is shown in figure 28.\*

*Application of elastohydrodynamic theory to practical problems*—The importance of the elastohydrodynamic film thickness, or of the normalized form obtained by division by some parameter representing the surface roughness, is now widely recognized. Its use as a criterion in the design of gears has been described by Seireg and Conry (ref. 62). Most of the

\* Cope, D. L.; and Haines, D. J.: The Direct Determination of Elastohydrodynamic Pressures in Gear Contacts by an Interferometric Technique. To be published by Inst. Mech. Engrs.

applications to gears have been limited to calculations of conditions at the pitch point. The film thickness does not vary much throughout the meshing cycle for gears with a speed ratio near unity, but the variation is considerably wider for gears with large speed ratios and should be taken into account when appropriate.

Complete rolling-contact bearings have been analysed by the use of the elastohydrodynamic theory. The paper by Wellons and Harris and the contributed discussion deal more fully with this aspect.

The importance of elastohydrodynamics in cam and tappet systems has been less widely recognized. Müller (ref. 63) reports a case in which the relative performance of two different cam designs conflicted with the rating by the usual criterion of maximum Hertzian stress on the nose contact, but was consistent with the theory of elastohydrodynamic lubrication. Figure 29 (ref. 64) shows the results of Naylor's calculations of film thicknesses for the two cams. Cam no. 1 suffered from severe scuffing, followed by pitting, but no. 2 was satisfactory.

The importance of elastohydrodynamic film thickness has been recognized in wear in rolling contact (ref. 65), wear and rippling in sliding contact (ref. 66), and fatigue in rolling contact (ref. 46). Snidle and Archard (ref. 67) have calculated film thicknesses in hypoid gears. Some of their results are shown in figure 30. Other things being equal, the film thickness increases with the offset between the axes of the pinion and of the gear, but other things will probably not be equal. As the offset is increased, the sliding speed, the heat generated by friction, and the gear bulk temperatures would all be expected to increase. Therefore, it is likely that the equilibrium film thickness will actually decrease with increasing offset at constant ambient temperature and under constant cooling conditions. The results of elastohydrodynamic theory would then be consistent with the observation that hypoid gears score more readily than do spiral bevel gears in which the offset is zero.

*Areas still to be studied*—Thermal effects have been introduced into the theories for the film shape and pressure distribution for a Newtonian lubricant, but the results have not contributed much to our understanding of elastohydrodynamic lubrication. No such theory gives a satisfactory account of the frictional traction results, and a combined thermal and viscoelastic solution to the complete elastohydrodynamic problem is needed. It will not be a simple matter to construct one. To accommodate dynamic problems in real machinery such as gears, cams, and tappets, etc., squeeze film effects should be included with the other ingredients.

Even with these additions, the theory would deal with ideally smooth surfaces, whereas practical engineering components have surfaces that are rough on a scale comparable with the minimum film thickness. Some way must therefore be found of incorporating real surfaces into the thermal, viscoelastic, squeeze, and wedge system.



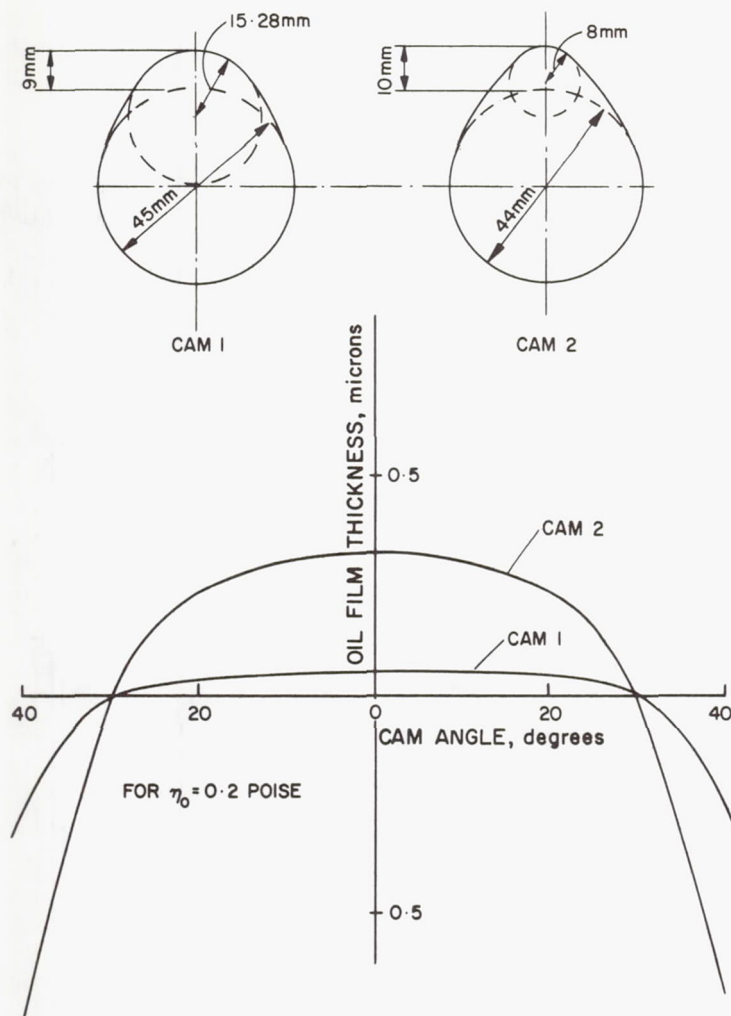


FIGURE 29.—Oil film thickness against degrees of cam angle (from details of Müller).

Finally work on plastrohydrodynamic lubrication, such as exists in metal-forming operations, is needed and should be of considerable help in solving practical problems.

*Acknowledgments*—Figures 14, 15, 17, 18 and 20 to 30 have been reproduced by permission of the Council of the Institution of Mechanical Engineers from the Proceedings of the Institution and from preprints of papers to be published in the Proceedings, and by permission of the individual authors named in the references. Figure 16 has been reproduced by permission of the Institute of Petroleum and Dr. A. W. Crook

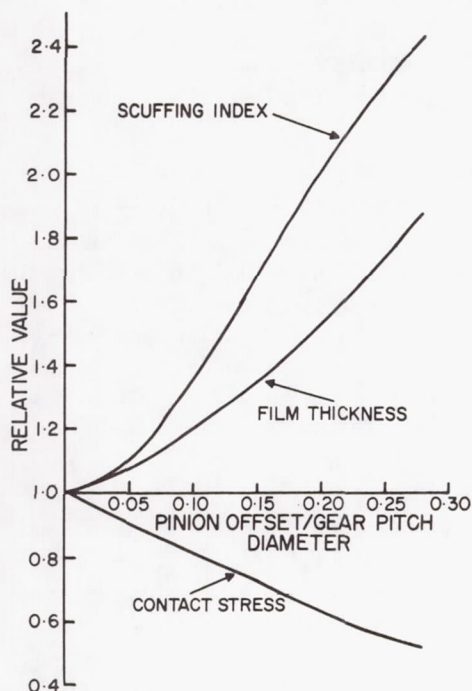


FIGURE 30.—Effect of hypoid pinion offset upon performance factors.

P. M. Ku (Southwest Research Institute, San Antonio, Texas)

Professor Dowson has given an authoritative account of the history and current understanding of elastohydrodynamics, to which he himself has made very notable contributions. The film profile and pressure profile for the one-dimensional case, with low sliding velocities and smooth surfaces, have shown generally satisfactory agreement between theory and experiments.

With regard to the film and pressure profiles, it appears that there are four areas in which further clarification would be helpful:

(1) How important is the role of the adsorption of the lubricant on the metal surfaces or on the oxide films on the surfaces? In other words, have instances been observed which show measured film thicknesses substantially different from the computed ones and more than can be accounted for by the bulk rheological considerations?

(2) The film profile for the two-dimensional case has been experimentally observed using the optical interference technique (ref. 68). A theoretical solution for the two-dimensional pressure profile is of interest to ball bearing applications.

(3) The effect of high sliding velocity, combined with rolling, on film and pressure profiles is of interest to gear lubrication. How important

are the viscous heating and the heat transfer effects, and how can they be quantitatively accounted for?

(4) What is likely to be the effect of surface roughness and topography on film and pressure profiles, and on the transition from EHD to boundary lubrication?

As pointed out by Professor Dowson, friction in sliding/rolling systems under EHD operation is difficult to interpret. Some early experiments were performed by Crook (ref. 24) but for very low loads. The coefficient of friction,  $f$ , computed from Crook's traction results, can be shown, for the range of higher sliding velocities, to decrease with an increase in (1) sliding velocity,  $V_s$ , (2) sum velocity,  $V_t$ , and (3) conjunction inlet viscosity,  $\mu_0$ ; and to increase with an increase in unit load,  $W$ . For a constant  $\mu_0$  ( $\sim 0.4p$ ), the relationship is given as a function of  $\mu_0 V_s/W$  in figure 31, which shows several distinct curves. Interestingly if the same results are plotted as a function of  $\mu_0 V_s V_t^2/W^2$ , then a single curve as shown in figure 32 is obtained.

For a wider range of variables, and for disks of greater surface roughness, Kelley obtained a correlation with  $\mu_0 V_s V_t^2/W$  for three straight mineral oils of different viscosity grades (ref. 69). He also obtained correlation with  $V_s V_t^2/\mu_0^7 W^{7.9}$  for a chlorinated methyl-phenyl silicone (ref. 70).

These are all presumably for the variation of  $f$  in the EHD regime. The parameters  $\mu_0 V_s V_t^2/W$ ,  $\mu_0 V_s V_t^2/W^2$ , and  $V_s V_t^2/\mu_0^7 W^{7.9}$  are not dimension-

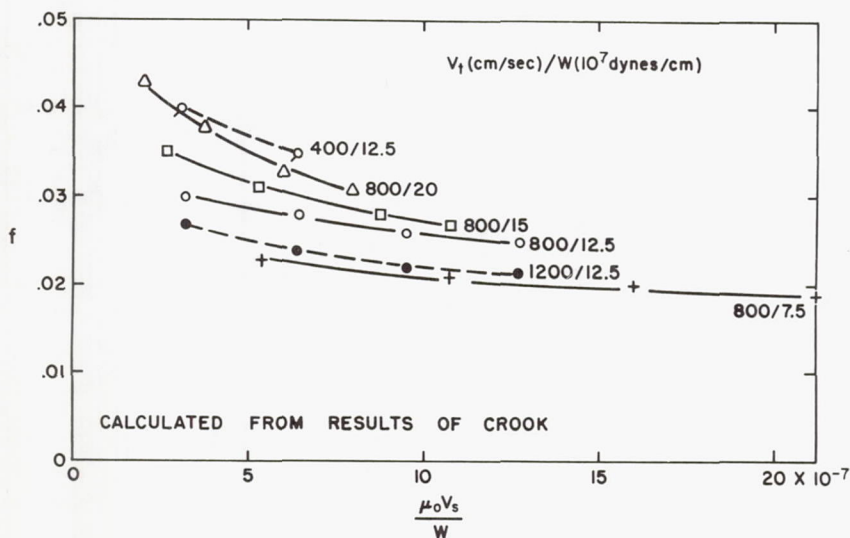


FIGURE 31.—Crook's friction data (ref. 24) plotted as a function of  $\frac{\mu_0 V_s}{W}$ .



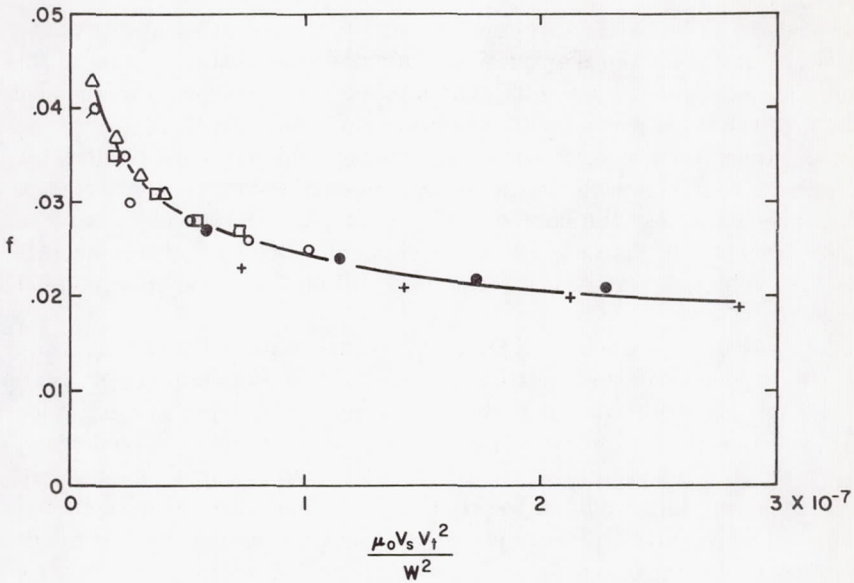


FIGURE 32.—Crook's friction data (ref. 24) plotted as a function of  $\frac{\mu_0 V_s V_t^2}{W^2}$ .

less, suggesting that something is missing. Nevertheless the question still remains as to how this type of behavior can be reconciled with that in the classical hydrodynamic regime for which  $\mu_0 V_s/W$  suffices to describe  $f$ , and with that in the boundary lubrication regime where the surface properties are known to be important? Why does  $f$  in the EHD regime depend upon the operating variables in the manners shown? To what extent and why do the surface properties, such as the composite surface roughness, enter into the picture? Is it possible that some form of asperity interaction, such as micro-elastohydrodynamics (ref. 71), is taking place? Is it possible that adsorption is rearing its ugly head? How can these effects, if significant, be accounted for quantitatively?

**C. A. Foord (Rolls-Royce Limited, Aero Engine Division, Derby, England)**

The (film thickness/ $\Sigma$  roughness) ratio is, as Professor Dowson points out, one of the most valuable parameters to emerge from EHD theory. Its use, however, poses the problem of how to sum the surface roughness, and what value to use for film thickness. These comments are confined to the film thickness problem.

The formula given for film thickness in point contact is close to the experimental formula

$$H = 1.37 \frac{U^{2/3} G^{.6}}{W^{.05}}$$

which was obtained from optical interference measurements (refs. 68 and 72) and is for the film thickness on the central plateau region of the contact, as shown in figure 33. The minimum film thickness in the side lobes of the contact is only 40 percent of this value. In this type of contact, capacitance measurements and theoretical methods predict the plateau film thickness. This applies over most of the contact, whereas an X-ray beam across the contact will be stopped by the side lobes and therefore tend to determine the minimum film thickness. Consequently  $(h/\Sigma)$  evaluated from X-ray predictions will be a much smaller number than the others.

The speed parameter,  $U$ , of figure 33 is of the same order as that found in gas turbine main shaft bearings, whereas the load parameter  $W'$  is an order of magnitude too small. Extrapolating by an order of magnitude from current results indicates that the side lobe thickness will be down to 17 percent of the central value at  $W' = 2.3 \times 10^{-5}$ . The side lobes will occupy a very small part of the contact area under these conditions but will still affect X-ray measurements; the film thickness at the rear of the contact is less capricious and remains at 70 to 75 percent of the control value. X-ray measurements are therefore likely to be more reliable if made in the direction of rolling.

In figure 34 the central and minimum film thicknesses are plotted against  $U$ . The minimum film thickness of zero in the side lobes at low

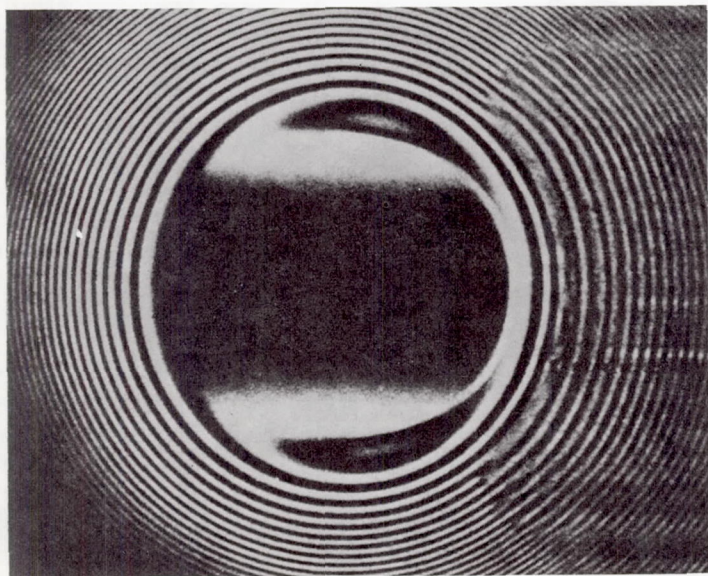


FIGURE 33.—Interference fringes of point contact.  $U = 2.32 \times 10^{-11}$ ,  
 $W' = 2.3 \times 10^{-6}$ ,  $G = 3200$ .

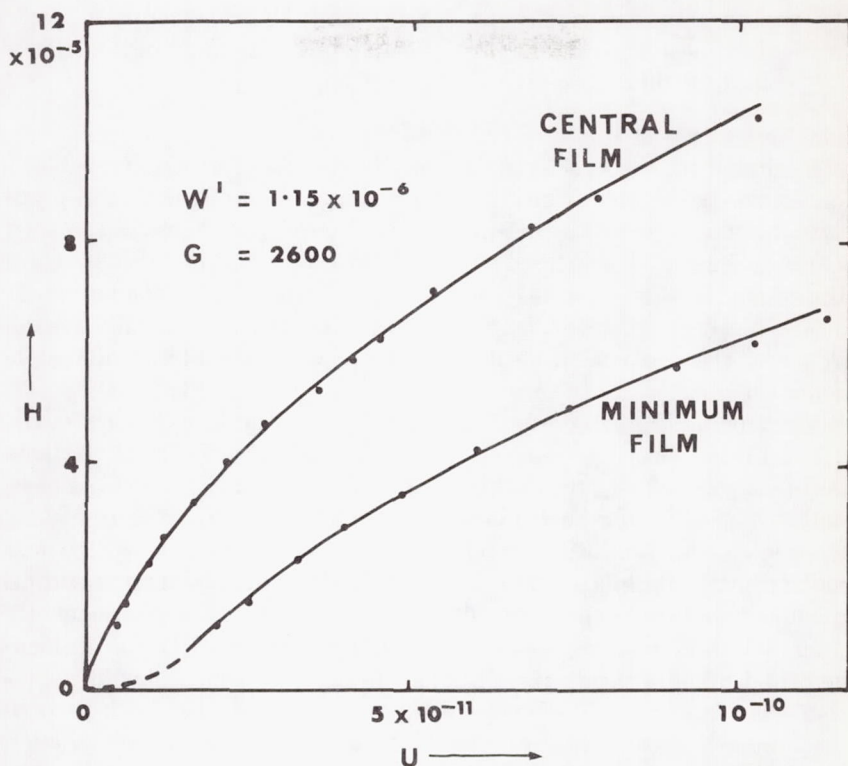


FIGURE 34.—Film thickness vs speed parameter.

speed is consistent with the pair of scratches observed after starting the optical measurement apparatus under load.

The main problem in using film thickness formulae is the lack of information on fluid properties. The results of references 68 and 72 show that relative effectiveness of fluids in forming elastohydrodynamic films can be expressed as a relative pressure-viscosity coefficient. The coefficients found in this way, however, range in value from  $0.64 \times 10^{-4} \text{ (psi)}^{-1}$  for isopropylbiphenyl to  $4.0 \times 10^{-4} \text{ (psi)}^{-1}$  for 5 phenyl-4 ether. This range of values is much greater than that found in low shear measurements and is probably due to non-Newtonian effects. Another difficulty is that film thickness formulae assume the inlet to be full of lubricant. In disc machines lubricated with heavy petroleum oils, sufficient lubricant may adhere to the disc surfaces; but with thin fluids in the optical measurement apparatus, this is not the case. The problem of oil starvation has been studied.\*

\* Data by L. D. Wedeven and A. Cameron to be published.



I hope that measurements will elucidate these problems in the near future, but meanwhile it will help if people quoting values of film thickness state how they were measured or calculated.

**R. S. Fein (Texaco Research Center, Beacon, New York)**

A number of speakers (e.g. Dyson, Harris, Ku, Naylor, and Sibley) have indicated the need for improved determination of rheological properties of lubricants at high pressures and shear rates. The indicated need seems to be based principally on measured friction tractions that are lower than theoretical values under nonisothermal EHD conditions. It seems, however, that the numerical disparity at least partially results from the thermal insulating characteristics of comparatively immobile boundary films.

My paper in this symposium\* points out that even nonreactive straight mineral oils seem to produce organometallic chemical reaction products that tend to remain at the load-bearing surfaces as films. Film thicknesses range generally from invisible ( $< \sim 4\mu$  inches) up to readily visible ( $> \sim 40\mu$  inches); the latter are varnish-like in appearance (ref. 73). These reaction-product films probably have thermal conductivities roughly comparable to bulk oil and about 100 times lower than steel. Consequently whenever the boundary film thickness approaches the EHD film thickness under nonisothermal conditions, the boundary film should appreciably insulate the EHD film from the steel. The result of such thermal insulation should be a decrease in EHD film viscosity and, hence, friction traction.

#### LECTURER'S CLOSURE

My disappointment at being unable to attend the NASA symposium on "Interdisciplinary Approach to the Lubrication of Concentrated Contacts" was mitigated by the knowledge that Mr. A. Dyson had undertaken to present my paper at short notice. The result is that delegates and readers of the proceedings will have two papers on elastohydrodynamics for the price of one. I would like to express my appreciation to Mr. Dyson and to Mr. P. M. Ku for dealing so effectively with my absence.

There are four discussions of my paper, but before considering them I would like to make a few comments on Mr. Dyson's contribution. Our reviews of the history and present state of the art appear to be broadly in agreement, but Mr. Dyson has noted one or two particularly worthy points.

It seems to be generally agreed that the failure of experimental film thickness to match the simple isothermal predictions of elastohydrodynamic theory at large values is attributable to heating of the lubricant

---

\* See R. S. Fein's lecture on Chemistry of Concentrated-Conjunction Lubrication.

by viscous dissipation in the inlet region. It seems surprising that this viewpoint has not been placed beyond dispute by a simple modification to the theory which allows for both viscous dissipation and adiabatic compression of the lubricant in the inlet region. The latter effect does not appear to be very important when the initial deviation between experimental and theoretical film thickness is noted; but since the large film thickness cases are usually encountered at very high speeds, it should not be dispensed with too readily.

Wilson's previously unpublished measurements of very thin films reported by Mr. Dyson add greatly to the store of knowledge on the question of limits to successful elastohydrodynamic action. It is remarkable that the empirical elastohydrodynamic equation based upon numerical solutions of the governing equations founded on continuum mechanics gives such good predictions of the film thickness when films of about  $100\text{\AA}$  are concerned. This is about four or five times the length of typical boundary lubricant molecules. Although Mr. Dyson does not report on the surface quality, it is unlikely that the c.l.a. will be less than the measured film thickness of  $100\text{\AA}$ . Does this imply a strong micro-elastohydrodynamic action?

Another very significant observation in relation to Mr. Dyson's report on very thin films is the statement that there is no trace of the long-range influences that many writers believe to be associated with boundary lubricant films. It would appear that measurements of film thickness under these extreme elastohydrodynamic conditions could throw some light on the multi-molecular layer and surface viscosity concepts that have emerged from boundary lubrication studies in recent years.

An important recent development in elastohydrodynamics reported by Mr. Dyson is the effect of high slide/roll ratios upon film thickness. The somewhat surprising ability of the contact to support large loads when the discs have equal and opposite velocities is accounted for in terms of the transverse temperature gradient resulting from large differences in disc surface temperatures. It is interesting to note that this may be the first successful demonstration of the principle outlined by Cameron (ref. 74) since the initial suggestion that it might account for parallel-surface thrust bearing performance was shown to be invalid (ref. 75). There is, however, another factor that should be explored in this case, and that is the differential thermal expansion of the disc surfaces. Professor Ling (ref. 76) studied the general problem some time ago, and the effect appeared to be quite potent as a load-generating mechanism. With equal and opposite surface velocities, it is possible that the symmetry of the situation will prevent load-carrying pressure generation due to thermal distortion. This will leave the field clear for a solution along the lines suggested by Mr. Dyson.

In his discussion of grease lubrication under elastohydrodynamic con-



ditions, Mr. Dyson refers to experiments under starved lubrication conditions. I believe that this touches on a field of activity that will be most important in elastohydrodynamic studies in the next few years. The early impressions from elastohydrodynamic theory suggested that the starting point of the pressure curve, or in practical terms the availability of lubricant, had little effect on minimum film thickness. However, it is now generally recognized that many elastohydrodynamic contacts operate under starved lubrication conditions and that this effect needs to be studied. The starvation may occur as a result of inadequate lubricant supply to a single contact, or it may result from the sweeping away of excess lubricant and the operation in a cavitated wake of a previous contact as in the rolling element bearing.

I am grateful to Mr. Ku for his interest and discussion of my paper. To some extent his first question has been answered by Mr. Dyson's report on the lack of evidence of long-range forces influencing film thickness measurements. I do not have any knowledge of direct measurements under elastohydrodynamic conditions which suggest that surface films are starting to influence contact behavior. I think that as far as surface roughness and boundary films are concerned, it is most surprising that their effects upon film thickness have not shown up in experiments concerned with theoretical separations of molecular and surface-irregularity proportions.

I certainly agree that a full theoretical solution for point contact pressure profiles is desirable, and it will no doubt be forthcoming in the future. The main problem is the enormity of the computational task. Work on cylindrical contacts in which side-leakage has been neglected has shown the importance of careful computation. It is not uncommon to consider intervals of one-thousandth of the Hertzian half-zone width. If similar increments are necessary when the field problem presented by point contacts is considered, it will require a major computational effort.

Little work has been done on the effect of high sliding velocity combined with rolling on film and pressure profiles, but the evidence that does exist (ref. 29) suggests that the film thickness can be predicted with fair accuracy under high sliding conditions if the viscosity appropriate to the inlet temperature is employed. This means that adequate analysis of gear meshing cycles depends on a successful measure of the overall thermal condition. I think that the time will soon be reached when a step-by-step analysis of a gear meshing cycle involving thermal and squeeze film effects will be undertaken.

Little can be said about the effect of surface roughness and topography on film and pressure profiles at the present time. The problem is difficult for both experimenters and theoreticians. One of the major problems is to establish a realistic model of the surfaces. It seems unlikely that the effect will be as great as a preliminary examination of the size of surface



irregularities in relation to the film thickness might suggest, since local or micro-elastohydrodynamic action will readily reduce the height of the asperities without producing a large transient perturbation of pressure. A useful indication of this effect was given by Higginson (ref. 37) in relation to squeeze film action. The persistence of coherent elastohydrodynamic films reported at the symposium by Dyson also suggests that micro-elastohydrodynamic action may delay surface interaction. On the other hand, evidence has been steadily accumulating that some form of asperity interaction occurs when the theoretical film thickness falls to a small multiple of the sum of the surface roughnesses.\* There will clearly be a mixed lubrication regime where both elastohydrodynamic and boundary lubrication actions combine to determine the overall contact characteristics. The writer has thought that this would occur when the ratio of calculated film thickness to surface roughness (sum of c.l.a.) was in the range of 1 to 5. The question of transition from elastohydrodynamic to boundary lubrication has been discussed by Dowson (ref. 43).

The correlation of Crook's friction coefficients with the single parameter  $\mu_0 V_s V_t^2 / W^2$  and Kelley's with  $\mu_0 V_s V_t^2 / W$  or  $V_s V_t^2 / \mu_0^7 W^{7.9}$  depending on lubricant type reported by Mr. Ku is most interesting. The different forms of the groups, none of which are dimensionless, suggest that a true representation of the complex phenomenon has yet to be found. If dimensionless groups are constructed, the well-known materials ( $G$ ), speed ( $U$ ), and load ( $W$ ) parameters of isothermal analysis will be encountered together with the ratio  $V_s/V_t$  and groups concerned with the thermal properties of the lubricant and solids. The main additional properties are thermal conductivity, specific heat, and density. A dimensionless group similar to that used by Ku arises from the following:

$$U = \frac{\eta_0 V_t}{E'R}$$

$$W = \frac{w}{E'R}$$

$$\bar{V} = \frac{V_s}{V_t}$$

These can be combined to give a new group  $\frac{\eta_0^2 V_s V_t}{w^2}$ . When Crook's data

for  $f$  is plotted against this new group, the original curves are condensed on to a single line within a similar accuracy to that portrayed by Ku's representation as shown in figure 35. The original curves can still be recog-

\* See F. W. Wellons and T. A. Harris' lecture on Bearing Design Considerations.

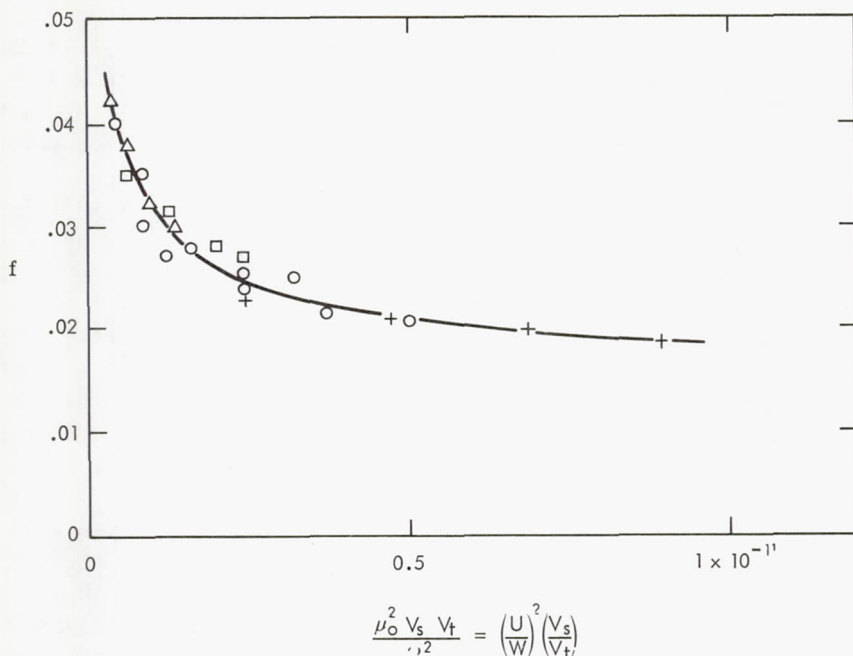


FIGURE 35.—Crook's friction data plotted against dimensionless elastohydrodynamic parameters.

nized. This suggests that some factor is still missing, but it would be interesting to see how Kelley's data responded to the same treatment.

Dr. Foord calls for a clear statement of the basis of assessment of both film thickness and surface quality in the formulation of the important ratio of film thickness/surface roughness. I agree with this plea, but I feel that we must not attempt to ascribe too high a degree of accuracy to the calculation of this ratio in the present state of knowledge. The evidence available from four ball and disc machines and life tests on rolling-contact bearings and gears shows that the ratio of calculated film thickness to surface roughness based on c.l.a. or rms values can be related directly to the degree of surface distress and to the life of the components. The fact that the film thickness in the lobes of spherical contacts is considerably lower than that calculated by the film thickness formula is presumably incorporated in these findings. An understanding of the mechanism of failure and the role of lubricant film thickness and surface quality will call for a more precise understanding of all the detailed features. At the present time, however, the overall effect is probably described with sufficient accuracy for the designer by the ratio as defined here.

Dr. Fein makes an interesting suggestion that might account for some of the observed discrepancies between measured and theoretical friction forces under elastohydrodynamic conditions. An insulating layer of the kind described by Dr. Fein should certainly lead to higher film temperatures, and it would be useful to have some quantitative measure of the effect.

Most of the contributions have discussed points concerned with the limits of successful elastohydrodynamic lubrication making it clear that there is much interest in the transition from elastohydrodynamic to boundary lubrication conditions and the onset of failure. I am grateful for the interest expressed by the contributors, and I hope that some of the unanswered questions will form subjects for future research.

#### REFERENCES

1. DOWSON, D.; AND HIGGINSON, G. R.: *Elastohydrodynamic Lubrication—The Fundamentals of Roller and Gear Lubrication*. Pergamon Press (London), 1966.
2. ANON.: *Elastohydrodynamic Lubrication*. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965–66.
3. DOWSON, D.: *Elastohydrodynamic Lubrication, an Introduction and Review of Theoretical Studies*. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965–66, p. 7.
4. MARTIN, H. M.: *Lubrication of Gear Teeth*. Engineering (London), vol. 102, 1916, p. 199.
5. PEPPLER, W.: *Untersuchungen uber die Druckubertragung bei belasteten und geschmierten unclaufenden achsparallelen Zylindern*. Maschinenelemente—Tagung (Aachen), 1935, VDI Verlag, Berlin.
6. PEPPLER, W.: *Druckubertragung und geschmierten zylindrischen Gleit- und Walzflächen*. VDI Forschung, vol. 391, 1938.
7. MELDAHL, A.: Contribution to the Theory of the Lubrication of Gears and of the Stressing of the Lubricated Flanks of Gear Teeth. Brown Boveri Review, vol. 28, no. 11, 1941, p. 374.
8. GATCOMBE, E. K.: *Lubrication Characteristics of Involute Spur Gears—A Theoretical Investigation*. Trans. ASME, vol. 67, 1945, p. 177.
9. BLOK, H.: *Fundamental Mechanical Aspects of Thin Film Lubrication*. Ann. N.Y. Acad. Sci., vol. 53, 1950, p. 779.
10. MCEWEN, E.: *The Effect of Variation of Viscosity with Pressure on the Load Carrying Capacity of Oil Films Between Gear Teeth*. J. Inst. Pet., vol. 38, 1952, p. 646.
11. MERRITT, H. E.: *Worm Gear Performance*. Proc. Inst. Mech. Engrs., vol. 129, 1935, p. 127.
12. GRUBIN, A. N.: *Central Scientific Research Institute for Technology and Mechanical Engineering*. Book 30, Moscow (D.S.I.R. Translation 337), 1949.
13. PETRUSEVICH, A. I.: *Fundamental Conclusions from the Contact—Hydrodynamic Theory of Lubrication*. Iz. Akad. Nauk. S.S.S.R. (OTN), vol. 2, 1951, p. 209.
14. DOWSON, D.; AND HIGGINSON, G. R.: *A Numerical Solution to the Elastohydrodynamic Problem*. J. Mech. Engrg. Sci., vol. 1, no. 1, 1959, p. 6.
15. DOWSON, D.; AND HIGGINSON, G. R.: *The Effect of Material Properties on the Lubrication of Elastic Rollers*. J. Mech. Engrg. Sci., vol. 2, no. 3, 1960, p. 188.
16. DOWSON, D.; HIGGINSON, G. R.; AND WHITAKER, A. V.: *Stress Distribution in*



- Lubricated Rolling Contacts. Proc. Inst. Mech. Engrs., Symp. Fatigue in Rolling Contact, 1963, paper 6, p. 66.
17. DOWSON, D.; AND HIGGINSON, G. R.: New Roller-Bearing Lubrication Formula. Engineering (London), vol. 192, 1961, p. 158.
  18. DOWSON, D.; HIGGINSON, G. R.; AND WHITAKER, A. V.: Elastohydrodynamic Lubrication—A Survey of Isothermal Solutions. J. Mech. Engrg. Sci., vol. 4, no. 2, 1962, p. 121.
  19. ARCHARD, G. D.; GAIR, F. C.; AND HIRST, W.: The Elastohydrodynamic Lubrication of Rollers. Proc. Roy. Soc. (London), vol. A262, 1961, p. 51.
  20. OSTERLE, J. F.; AND STEPHENSON, R. R.: A Direct Solution of the Elastohydrodynamic Lubrication Problem. ASLE Trans., vol. 5, 1962, p. 365.
  21. CROOK, A. W.: The Lubrication of Roller, I. Phil. Trans. (London), vol. A250, 1958, p. 387.
  22. CROOK, A. W.: The Lubrication of Rollers, II. Film Thickness with Relation to Viscosity and Speed. Phil. Trans. (London), vol. A254, 1961, p. 223.
  23. CROOK, A. W.: The Lubrication of Rollers, III. A Theoretical Discussion of Friction and the Temperatures in the Oil Film. Phil. Trans. (London), vol. A254, 1961, p. 237.
  24. CROOK, A. W.: The Lubrication of Rollers, IV. Measurements of Friction and Effective Viscosity. Phil. Trans. (London), vol. A255, 1963, p. 281.
  25. SIBLEY, L. B.; AND ORCUTT, F. K.: Elastohydrodynamic Lubrication of Rolling Contact Surfaces. ASLE Trans., vol. 4, 1961, p. 234.
  26. DYSON, A.; NAYLOR, H.; AND WILSON, A. R.: The Measurement of Oil Film Thickness in Elastohydrodynamic Contacts. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965, p. 119.
  27. CAMERON, A.; AND GOHAR, R.: Theoretical Experimental Studies of the Oil Film in Lubricated Point Contact. Proc. Roy. Soc. (London), vol. A291, 1966, p. 520.
  28. CHENG, H. S.; AND STERNLICHT, B.: A Numerical Solution for the Pressure, Temperature and Film Thickness Between Two Infinitely Long, Lubricated Rolling and Sliding Cylinders, Under Heavy Loads. Trans. ASME, J. Basic Engrg., vol. 87D, 1965, p. 695.
  29. DOWSON, D.; AND WHITAKER, A. V.: A Numerical Procedure for the Solution of the Elastohydrodynamic Problem of Rolling and Sliding Contacts Lubricated by a Newtonian Fluid. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965, p. 57.
  30. DYSON, A.: Film Thickness in Elastohydrodynamic Lubrication by Silicone Fluids. Proc. Inst. Mech. Engrs., vol. 180, pt. 3K, 1966, p. 97.
  31. ARCHARD, J. F.; AND KIRK, M. T.: Lubrication at Point Contacts. Proc. Roy. Soc. (London), vol. 261, 1961, p. 532.
  32. ARCHARD, J. F.; AND COWKING, E. W.: Elastohydrodynamic Lubrication at Point Contacts. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1966, p. 47.
  33. CHRISTENSEN, H.: The Oil Film in a Closing Gap. Proc. Roy. Soc. (London), vol. A266, 1962, p. 312.
  34. DOWSON, D.; AND JONES, D. A.: Lubricant Entrapment Between Approaching Elastic Solids. Nature (London), 1967.
  35. BURTON, R. A.: Effects of Two-Dimensional, Sinusoidal Roughness on the Load Support Characteristics of a Lubricant Film. Trans. ASME, J. Basic Engrg., vol. 85D, 1963, p. 258.
  36. HAMILTON, D. B.; WALOWIT, J. A.; AND ALLEN, C. M.: A Theory of Lubrication by Micro Irregularities. Trans. ASME, J. Basic Engrg., vol. 88D, 1966, p. 177.
  37. HIGGINSON, G. R.: Discussion of Paper 12. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965-66, p. 243.

38. DOWSON, D.; AND WHOMES, T. L.: Effect of Surface Quality Upon the Traction Characteristics of Lubricated Cylindrical Contacts. *Proc. Inst. Mech. Engrs.*, vol. 182, pt. 1, 1968.
39. MOES, H.: Discussion of Paper R1. *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1966.
40. Viscosity and Density of Over Forty Lubricating Fluids of Known Composition at Pressures to 150,000 lb/in.<sup>2</sup> Gauge and Temperatures to 425° F. ASME Res. Comm. on Lub., 1953.
41. LAMB, J.: Viscoelastic Behaviour and the Lubricating Properties of Liquids. ASME Appl. Mech.—Fluids Engrg. Conf., Washington, 1965.
42. KANNEL, J. W.: The Measurement of Pressure in Rolling Contacts. *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1966.
43. DOWSON, D.: Transition to Boundary Lubrication from Elastohydrodynamic Lubrication. *Boundary Lubrication: An Appraisal of World Literature*, ASME Res. Comm. on Lub., 1969, p. 229.
44. DOWSON, D.: Die elastohydrodynamische Schmierung von Wälzlagern. *Maschinenbautechnik (Schmierungstechnik)*, 1969, p. 106.
45. DOWSON, D.; AND HIGGINSON, G. R.: A Theory of Involute Gear Lubrication. *Inst. Pet. Gear Lub. Symp.* 1964.
46. DAWSON, P. H.: Effect of Metallic Contact on the Pitting of Lubricated Rolling Surfaces. *J. Mech. Engrg. Sci.*, vol. 4, no. 1, p. 16.
47. DAWSON, P. H.: The Effect of Metallic Contact and Sliding on the Shape of the S-N Curve for Pitting Fatigue. *Inst. Mech. Engrs. Symp. on Fatigue in Rolling Contacts*, 1963, paper 4, p. 41.
48. TALLIAN, T. E.; MCCOOL, J. J.; AND SIBLEY, L. B.: Partial Elastohydrodynamic Lubrication in Rolling Contact. *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1965-66, p. 169.
49. MARTIN, H. M.: *Lubrication of Gear Teeth*. Engineering (London), vol. 102, 1916, p. 199.
50. GRUBIN, A. N.: Fundamentals of the Hydrodynamic Theory of Heavily Loaded Cylindrical Surfaces. Book 30, Moscow (D.S.I.R. Translation 337), 1949.
51. PETRUSEVITCH, A. I.: Fundamental Conclusions from the Contact—Hydrodynamic Theory of Lubrication. *Iz. Akad. Nauk. SSSR (OTN)*, vol. 2, 1959—p. 209.
52. CROOK, A. W.: Developments in Elastohydrodynamic Lubrication. *J. Inst. Pet.*, vol. 49, 1963, p. 295.
53. DYSON, A.: Investigation of the Discharge Voltage Method of Measuring the Thickness of Oil Films Formed in a Disc Machine Under Conditions of Elastohydrodynamic Lubrication. *Proc. Inst. Mech. Engrs.*, vol. 181, pt. 1, 1966-67, p. 633.
54. DYSON, A.; AND WILSON, A. R.: Film Thicknesses in Elastohydrodynamics Lubrication at High Slide/Roll Ratio. *Inst. Mech. Engrs. Tribology Conv.* Gothenburg, Sweden, May 1969, paper 11.
55. CAMERON, A.: Hydrodynamic Lubrication of Rotating Discs in Pure Sliding. A New Type of Oil Film Formation. *J. Inst. Pet.*, vol. 37, 1951, p. 471.
56. BARLOW, A. J.; HARRISON, G.; AND LAMB, J.: Viscoelastic Relaxation of Polydimethyl Siloxane Liquids. *Proc. Roy. Soc. (London)*, vol. A282, 1964, p. 228.
57. DYSON, A.; AND WILSON, A. R.: Film Thickness in Elastohydrodynamic Lubrication by Silicone Fluids. *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3K, 1965-66, p. 97.
58. DYSON, A.; AND WILSON, A. R.: Film Thicknesses in Elastohydrodynamic Lubrication of Rollers by Greases. *Inst. Mech. Engrs. Symp. on Grease Lubrication*, London, England, February 19-20, 1970, paper 18.

59. THEYSE, F. H.: Elastohydrodynamik. Schmiertechnik, vol. 15, no. 1, 1958, p. 15.
60. GREENWOOD, J. A.: Presentation of Elastohydrodynamic Film Thickness Results. J. Mech. Engrg. Sci., vol. 11, no. 2, 1969, p. 128.
61. HERREBRUGH, K.: Solving the Incompressible and Isothermal Problem in Elastohydrodynamic Lubrication Through an Integral Equation. Trans. ASME, J. Lub. Tech., vol. 90F, 1968, p. 262.
62. SEIREG, A.; AND CONRY, T.: Optimum Design of Gear Systems for Surface Durability. ASLE Trans., vol. 11, 1968, p. 321.
63. MÜLLER, R.: The Effect of Lubrication on Cam and Tappet Performance. M.T.Z., vol. 27, 1966, p. 58.
64. NAYLOR, H.: Cams and Friction Drives. Inst. Mech. Engrs. Interlat. Conf. on Lubrication and Wear, London, England, Oct. 1967, paper 15.
65. TALLIAN, T. E.; MCCOOL, J. I.; AND SIBLEY, L. B.: Partial Elastohydrodynamic Lubrication in Rolling Contact. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965-66, p. 169.
66. LANDEN, E. W.: Slow Speed Wear of Steel Surfaces Lubricated by Thin Oil Films. ASLE Trans., vol. 11, 1968, p. 6.
67. SNIDLE, R. W.; AND ARCHARD, J. F.: Lubrication at Elliptical Contacts. Inst. Mech. Engrs. Tribology Conv., Gothenburg, Sweden, May 1969, paper 17.
68. FOORD, C. A.; HAMMANN, W. C.; AND CAMERON, A.: Evaluation of Lubricants Using Optical Elastohydrodynamics. ASLE Trans., vol. 11, 1968, p. 31.
69. BENEDICT, G. H.; AND KELLEY, B. W.: Instantaneous Coefficients of Gear Tooth Friction. ASLE Trans., vol. 4, 1961, p. 59.
70. LEACH, E. F.; AND KELLEY, B. W.: Temperature—The Key to Lubricant Capacity. ASLE Trans., vol. 8, 1965, p. 271.
71. FEIN, R. S.; AND KREUTZ, K. L.: Discussion on Boundary Lubrication. Interdisciplinary Approach to Friction and Wear, NASA SP-181, 1968, p. 358.
72. FOORD, C. A.: Doctoral Thesis, Univ. of London, 1968.
73. FEIN, R. S.: Operating Procedure Effect on Critical Temperature. ASLE Trans., vol. 10, 1967, p. 373.
74. CAMERON, A.: Hydrodynamic Lubrication of Rotating Discs in Pure Sliding. A New Type of Oil Film Formation. J. Inst. Pet., vol. 37, 1951, p. 37.
75. DOWSON, D.; AND HUDSON, J. D.: Thermo-Hydrodynamic Analysis of the Infinite Slider Bearing. II The Parallel-Surface Bearing. Proc. Lub. and Wear Conv., Inst. Mech. Engrs., 1964, p. 45.
76. LING, F. F.; AND MOW, V. C.: Surface Displacement of a Convective Elastic Half-Space Under an Arbitrarily Distributed Fast-Moving Heat Source. Trans. ASME, J. Basic Engrg., vol. 87D, 1965, p. 729.



# Microslip and Creep in Contacts

H. PORITSKY

Consultant

Schenectady, New York

The first four sections contain a detailed discussion of microslip and creep phenomena for stationary and rolling contacts of cylindrical bodies. Emphasis is laid on the method of conformal mapping in the solution for the pertinent harmonic functions. Microslip and creep over Hertzian contact ellipses for noncylindrical bodies are reviewed in the fifth through seventh sections. Due to limitations of space and time, their treatment is not as thorough, and no attempt has been made to present the results of various authors in a unique notation and system of sign conventions. The fifth section is devoted to a review of some of the approximate analytical solutions and experimental measurements. The sixth section reviews the analytical solutions of Hertz and Mindlin emphasizing the analogies to classical problems in potential theory, as well as the more recent analytical solutions by de Pater, Kalker, and Mow et al. It also contains a brief introduction to ellipsoidal harmonics. Numerical methods of solution of the contact problems are considered in the seventh section. Thus far their success has been rather limited, but they hold great promise for the future. The method of integral equations for the tractions over the contact area look especially attractive, once the singularities of their integrands can be tamed. The final section reviews briefly the criticism by Nayak and Paul of the "smooth surface, constant coefficient of friction" theory which underlies existing theory of microslip and creep. It then discusses the effect of lubricants, reviews a paper by Johnson and Cameron, and concludes with mentioning some unsolved problems.

## 1. INTRODUCTION

**M**ICROSLIP IN CONTACT PHENOMENA means slip that may occur over the area of contact of two solids, even when there is no gross sliding between them, as in tractive rolling of a locomotive wheel on a track (refs. 1 to 3). As an illustration, we recall the familiar theory of tractive rolling of cylinders  $C_1$ ,  $C_2$  or of a cylinder on a plane when  $C_1$ ,  $C_2$  are elastically similar, i.e., have the same elastic constants.

The Hertz contact theory shows that contact occurs over a strip of width  $2a$ , given by

$$a^2 = \frac{2P}{\pi} \left( \frac{1-\nu_1}{G_1} + \frac{1-\nu_2}{G_2} \right) / \left( \frac{1}{R_1} + \frac{1}{R_2} \right) \quad (1)$$

where  $P$  = compressive force between the cylinders, per unit axial length,

$R_1, R_2$  = radii of  $C_1, C_2$ ,

$G_1, G_2$  = shear (or rigidity) moduli of  $C_1, C_2$ ,

$\nu_1, \nu_2$  = Poisson ratios of  $C_1, C_2$ .

The pressure distribution over the contact zone is

$$p(x) = \frac{2P}{a^2} \sqrt{a^2 - x^2}, \quad |x| < a \quad (2)$$

where  $x$  is the distance from the zone center line.

Figure 1 shows a locomotive driving wheel, rolling to the right on a level track with velocity  $U$ . In addition to the normal force  $P$ , the wheel pushes on the track with a tractive force  $-Q$  per unit  $z$ . The reactions of the track on the driving wheels furnish the force necessary to drive the train (fig. 2).

The force  $Q$  is distributed as a shear force  $q(x)$  over the contact strip.  $Q$  can never exceed the value  $\mu P$ , where  $\mu$  is the coefficient of friction (assumed constant):

$$0 < Q < \mu P = Q_m \quad (3)$$

When  $Q$  reaches its limiting value, then

$$q(x) = \mu P(x), \quad -a < x < a \quad (4)$$

and the wheel slides over the track.

We restrict  $Q$  as in eq. (3) and assume that the wheel and track have the same elastic constants:  $G_1 = G_2$ ,  $\nu_1 = \nu_2$ . Then eq. (1) yields

$$a^2 = \frac{4PR_1(1-\nu_1)}{\pi G_1} = \frac{8PR_1(1-\nu_1^2)}{\pi E_1} \quad (5)$$

For  $\nu = .3$ , this can be put in the form

$$a/R_1 = 1.78\bar{p}/G_1 \quad (6)$$

where  $\bar{p}$  is the average pressure over the contact zone. The ratio  $\bar{p}/G$  is generally quite small, hence,  $a/R_1$  is likewise small.

The small size of  $a/R_1$  is used to justify neglecting the curvature of the wheel beyond the contact zone, therefore assuming that the applied stresses  $p(x)$ ,  $q(x)$  spread away from the contact as in a semi-infinite solid.

When eq. (3) holds and no gross sliding takes place, it turns out that the contact zone is divided into two parts: (1) a forward part  $L$  extending over

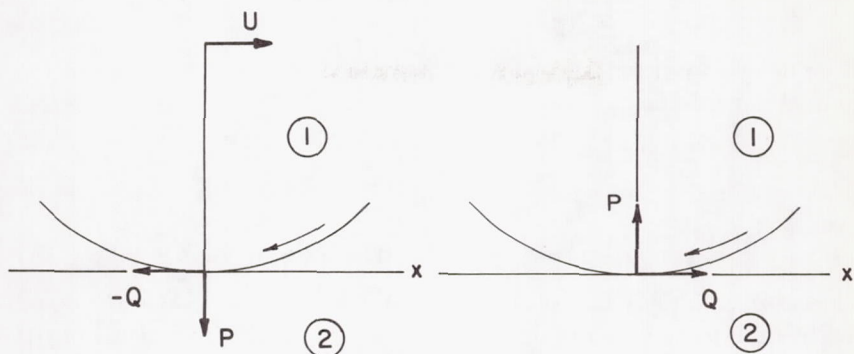


FIGURE 1.—Forces exerted by wheel on track. FIGURE 2.—Reaction forces of track on wheel.

$$L: \quad a - 2b < x < a \quad (7)$$

where

$$\frac{a}{b} = \sqrt{1 - \frac{Q}{\mu P}} = \sqrt{1 - \frac{Q}{Q_m}} \quad (8)$$

over which

$$q(x) < \mu p(x) \quad (9)$$

and over which no slipping takes place between the wheel and the track; and (2) the remaining portion  $S$

$$S: \quad -a < x < a - 2b \quad (10)$$

over which

$$q(x) = \mu p(x) \quad (11)$$

and over which slipping does occur. The rate of this slipping is small compared to the train velocity  $U$ ; hence, it is known as microslip. The notations  $L$  and  $S$  for the two parts of the zone of contact are used to bring out the notions of locked area and slip area, respectively.

The explanation of these features lies in the fact that, as a result of the action of the tractive force  $-q(x)$ , the track is stretched ahead of the zone of contact, compressed behind it, so that the tangential strain

$e_{x2} = \frac{\partial u_2}{\partial x}$  satisfies the inequalities:

$$e_{x2}(x) > 0 \text{ for } x > a, \quad e_{x2}(x) < 0 \text{ for } x < -a \quad (12)$$

Under the action of the shear  $q(x)$ , the wheel is compressed ahead of the contact zone and stretched behind it. In fact, outside the contact zone the wheel strain is opposite to that of the track:



$$e_{x1}(x) = -e_{x2}(x), \quad |x| > a \quad (13)$$

If we denote by  $e_1, e_2$  the strains in the wheel and track just ahead of the contact zone, then

$$e_1 = e_{x1}(a) < 0, \quad e_2 = e_{x2}(a) = -e_1 > 0 \quad (14)$$

It turns out that, in addition to eq. (9), the following relation holds over  $L$ :

$$e_{x1}(x) - e_{x2}(x) = e_1 - e_2 = 2e_1 < 0 \text{ for } (a - 2b) < x < a \quad (15)$$

That is, over  $L$  there is a *constant strain difference* between the wheel and the track, equal to the difference between these strains just ahead of the contact zone.

If we suppose that eq. (13) holds over  $L$  as well, eq. (15) would yield over  $L$

$$e_{x1}(x) = e_1, \quad e_{x2}(x) = e_2 = -e_1 \quad (16)$$

That is, the wheel portion of  $L$ , compressed in the ratio  $(1+e_2):1$ , would be "paid out" on the track portion of  $L$ , expanded in the ratio  $(1+e_2):1 \sim (1-e_1):1$ . In other words, the motion of the wheel would be the same as that of a wheel of radius

$$R_1(1+e_1) < R_1 \quad (17)$$

free from stress, rolled (without slipping) on a track, expanded in the ratio

$$(1+e_2):1 = (1-e_1):1 \quad (18)$$

Actually it will be shown presently that neither (13) nor (16) holds over  $L$ . However, they do apply to the component of strain  $e_{qx}$  produced by the tractive shear forces  $\pm q(x)$ . In addition to this strain component, there is a component of strain  $e_{xp}$  over the contact strip, induced by  $p(x)$ , and proportional to  $p(x)$ . This strain component is the same for the wheel and the track; therefore it does not produce any slip between them, and the above description of the relative motion turns out to be correct. Although the wheel and the track portions of  $L$  are additionally compressed to a strain pattern  $e_{xp}$  due to  $p(x)$ , this pattern is the same for the wheel and the track and does not interfere with the slip caused by the strain  $e_{xq}$  due to  $\pm q(x)$ , nor does it produce any slip of its own.

Over  $S$ , the remaining portion of the contact zone, eq. (11) holds, and there is slip. Again there is a strain component due to  $p(x)$ , proportional to it, which is the same for the wheel and the track and causes no slip. There is also a strain component due to  $\pm q(x)$ , equal and opposite for the wheel and the track. For the wheel it changes from  $e_1$  at  $x = a - 2b$ , to  $e_{x1}(-a) > e_1$  as  $x$  varies from  $x = a - 2b$  to  $x = -a$  (fig. 3).

As a result of the dimensional changes over  $L$  described above, the angular velocity of the wheel is

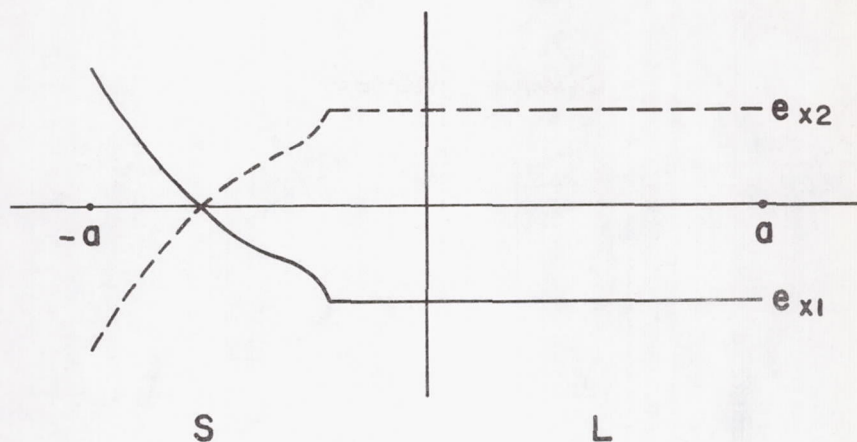


FIGURE 3.—Extensions of wheel surface (solid line) and of track (dashed line) over contact zone.

$$\omega = \omega_0(1 + e_2 - e_1) = \omega_0(1 - 2e_1) \quad (19)$$

where  $\omega_0 = U/R$  and  $U$  is the velocity of the train (or wheel axle center). The wheel appears to creep over the track with a velocity

$$U(e_2 - e_1) = -2e_1U = \xi U \quad (20)$$

where  $\xi$  is known as the creep ratio.

As rolling proceeds, figure 3 moves to the right relative to fixed co-ordinates, with velocity  $U$ , and the slip rate of the wheel over the track is

$$\frac{d}{d\ell}(e_{x1} - e_{x2}) = -U \left[ \frac{d(e_{xq})_1}{dx} - \frac{d(e_{xq})_2}{dx} \right] = -2U \frac{d(e_{xq})_1}{dx} \quad (21)$$

As a result of this slipping against the shear friction force  $q(x)$ , work is being done at the rate

$$W = -2U \int_{-a}^{a-2b} q(x) \frac{d(e_{xq})_1}{dx} dx \quad (22)$$

per unit  $z$ . This work is spent in wear and fretting of the wheel and track material.

A more detailed discussion of tractive rolling of elastically similar cylinders is given in the third section. The distributions  $p(x)$ ,  $q(x)$  are indicated in figure 4, and in equation form are given by eq. (2) and by

$$q(x) = \frac{2\mu P}{\pi a^2} \begin{cases} \sqrt{a^2 - x^2} & \text{over } S \\ \sqrt{a^2 - x^2} - \sqrt{b^2 - (x - a + b)^2} & \text{over } L \end{cases} \quad (23)$$

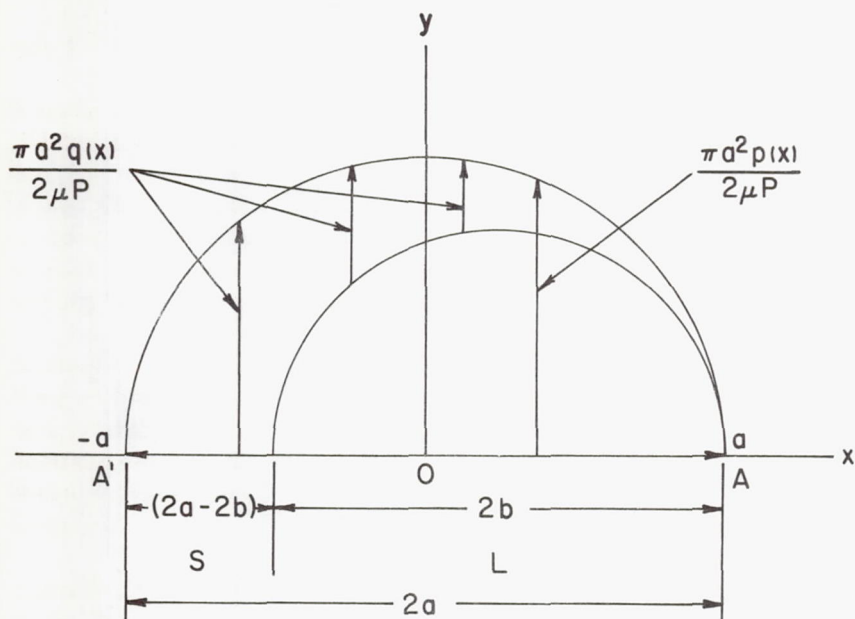


FIGURE 4.—Pressure and shear friction distributions over locked and slip portions of contact zone.

The strain components are given by

$$(e_{xp})_1 = (e_{xp})_2 = -\frac{(1-\nu_1)P}{\pi G_1 a}, \quad |x| < a \quad (24)$$

$$(e_{xq})_1 = -(e_{xq})_2 = \frac{2(1-\nu_1)P\mu}{\pi G_1 a} \begin{cases} -\left(1-\frac{b}{a}\right) & \text{in } L \\ -\left(1-\frac{b}{a}\right) + \sqrt{\left(1-\frac{x}{a}\right)\left(1-\frac{2b}{a}-\frac{x}{a}\right)} & \text{in } S \end{cases} \quad (25)$$

and schematic figure 3 can now be correctly completed; in particular, it has an infinite slope at  $x = a - 2b$ .

From eq. (25) there results

$$e_1 = (e_{xq})_1, \text{ in } L = -\frac{2(1-\nu_1)P\mu}{\pi G_1 a} \left(1-\frac{b}{a}\right)$$

$$(e_{xq})_1(-a) = \frac{2(1-\nu_1)P\mu}{\pi G_1 a} \left[ -\left(1-\frac{b}{a}\right) + 2\sqrt{1-\frac{b}{a}} \right] > 0 \quad (26)$$



Substitution into eqs. (20), (22) completes the calculation of the slip and losses.

Many workers in the field of rolling contacts (refs. 4 and 5) adopt a system of axes in which the state of stress is fixed. Thus in figures 1 and 2, such a system would have the contact strip fixed, and the wheel and track flow through it in direction of negative  $x$ , with velocities  $-U(1+\xi_1)$ ,  $-U(1+\xi_2)$ , respectively, at the origin. By reversing the train direction,  $-U$  above is changed to  $U$ , and the forward side of the contact zone is to the left. The advantage of such a procedure is that one may use the Euler (hydrodynamic) equations for describing the motion.

The second section reviews the general relation between the normal pressure  $p(x)$  and the shear traction  $q(x)$  applied at the boundary  $y=0$  of a semi-infinite solid and the displacements along the boundary, and applies the results to the contact problem of two cylinders. These relations are expressed in terms of two harmonic functions and their conjugate harmonics. In particular, the Hertz solution is derived, and the tangential components accompanying it are obtained.

The third section applies the results of the second section to the contact problem of elastically similar cylinders with both normal and shear tractions, and the derivation of the above equations is indicated. Also discussed in the third section is the case when the friction coefficient is infinite and slip over the contact interval is prevented, as well as contact of cylinders with a nonparabolic profile.

Contacts of elastically dissimilar cylinders are considered in the fourth section. Of special interest are the results of Bental and Johnson (ref. 6), obtained by an approximate numerical solution of the integral equations for  $p(x)$ ,  $q(x)$  expressing the conditions of slip and creep over the contact regions,  $S_1$ ,  $S_2$ ,  $S_3$  with the direction of slip being opposite in sequential area; these are separated by locked areas.

Slip and locked areas also occur in both stationary and rolling contacts of noncylindrical bodies, transmitting normal and tangential tractions across the elliptic Hertzian contact area, as well as a moment about the normal axis. For stationary contacts, slipping occurs when the forces are first applied or when they are varied in time. For rolling contacts, creep may occur over the locked areas in the direction of rolling and at right angles to it and may also involve spin about the normal axis.

An excellent review of microslip contact phenomena is given by Johnson (ref. 4), where a review of the state of the art through 1960 is given.

Compared to the case of cylindrical contacts where an exact solution for slip and creep phenomena is available, at least for the case of elastically similar bodies, very few exact analytical solutions exist for three-dimensional contacts. The fifth section reviews a number of approximate solutions of these problems and experimental measurements of the slip and creep (including the spin), as well as photoelastic determinations of the tractions over the contact ellipse.

The sixth section contains a review of some of the exact analytic solutions, while the seventh section presents some numerical solutions that have been carried out in trying to harness the use of modern computers to the numerical solution of the difficult analytical problems presented by three-dimensional contacts.

The concluding section discusses briefly the effect of lubrication of rolling contacts as related to the phenomena of creep and slip. It also presents some of the recent criticisms of the smooth surface assumption underlying existing theory. It is claimed that only a statistical representation of the departure of actual surfaces from the ideal smooth surfaces of the theory can explain the wide variations in the experimental results obtained by many investigators.

Creep phenomena over contacts result in wear and fretting of the contacting surfaces and, in some cases, cause failures of bearings. In jet engines, this friction of ball bearings, though small compared to the loads and power transmitted, may develop so much heat that its dissipation becomes an important design factor.

## 2. FORMULATION OF CONTACTS BETWEEN CYLINDERS IN TERMS OF HARMONIC FUNCTIONS; CONFORMAL MAPPING

The displacement vector and the stress and strain tensors for solutions of the linear elasticity equations in regions free from body forces satisfy the differential equation

$$\nabla^4 = 0, \quad \nabla^4 = \nabla^2(\nabla^2) \quad (27)$$

known as the repeated Laplace equation. The scalar solutions of this equation can be expressed in terms of two harmonic functions, i.e., two solutions of the Laplace equation

$$\nabla^2 = 0 \quad (28)$$

Hence, analytical solutions of elastic problems reduce to the determination of harmonic functions which, properly combined, satisfy given boundary conditions.

In two-dimensional problems of plane strain, where the components of the displacement vector and of the stress and strain tensors are independent of  $z$ , the Laplace operator in eqs. (27) and (28) reduces to  $\partial^2/(\partial x)^2 + \partial^2/(\partial y)^2$ . It is known that two-dimensional harmonic functions  $U(x, y)$  are related to analytic functions of the complex variable  $Z = x + iy$ , where  $i = \sqrt{-1}$ . Indeed, if  $w$  is an analytic function of  $Z$ :

$$w = f(Z) = U + iV \quad (29)$$

then  $U$ , the real part of  $w$ , is harmonic, and so is  $V$ , the imaginary part of  $w$ . The function  $V$  is known as the conjugate harmonic of  $U$ ; jointly they satisfy the Cauchy-Riemann equations. Conversely given any har-

monic function  $U(x, y)$ , there exists a conjugate harmonic function  $V(x, y)$ , determined except for an additive constant, such that eq. (29) holds.

The analytic functions connected with contact problems of cylinders were introduced by Carter (ref. 1). Referring to figure 5, for the upper half plane  $y > 0$ , loaded over the interval

$$y=0, \quad -a < x < a \quad (30)$$

with a normal pressure force  $p(x)$  and a shear force  $q(x)$ , we introduce, following Carter, the two analytic functions  $f(Z)$  and  $g(Z)$ , defined by

$$f(Z) = F(x, y) + iF'(x, y) = \int_{-a}^a p(s) \ln[(Z-s)/i] ds$$

$$g(Z) = G(x, y) + iG'(x, y) = \int_{-a}^a q(s) \ln[(Z-s)/i] ds \quad (31)$$

Equating real and imaginary parts

$$F = \int_{-a}^a p(s) \ln r ds, \quad F' = \int_{-a}^a p(s) \left( \theta - \frac{\pi}{2} \right) ds$$

$$G = \int_{-a}^a q(s) \ln r ds, \quad G' = \int_{-a}^a q(s) \left( \theta - \frac{\pi}{2} \right) ds \quad (32)$$

where

$$r = |Z - s|, \quad \theta = \arg(Z - s) = \tan^{-1}[y/(x-s)], \quad 0 \leq \theta \leq \pi \quad (33)$$

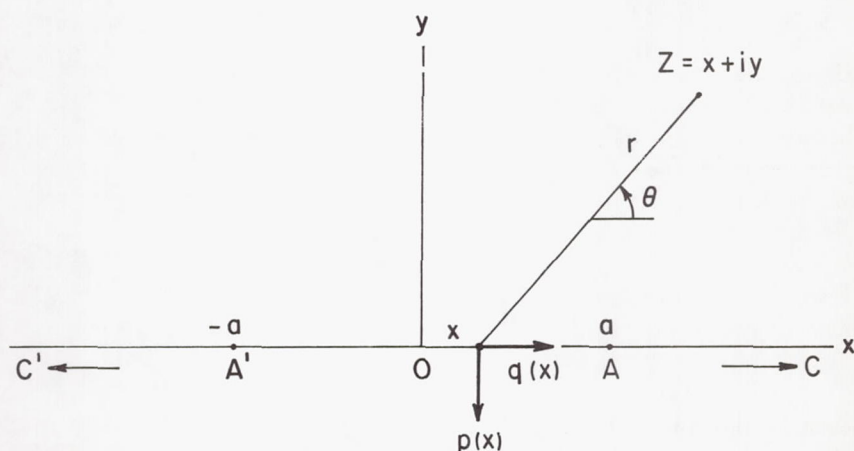


FIGURE 5.—Illustration for equations (31) to (33).



If the normal pressure load reduces to a concentrated normal force  $N$  at the origin, eqs. (31) and (32) simplify to

$$\begin{aligned} g &= 0, & f &= \ell n(Z/i) \\ F &= N \ell n r, & F' &= N \left( \theta - \frac{\pi}{2} \right) \end{aligned} \quad (34)$$

where  $r, \theta$  are polar coordinates (fig. 6). It can be shown now (ref. 2) that, along the boundary, the displacements are given by

$$\begin{aligned} n_1 &= -\frac{N(1-\nu_1)}{\pi G_1} \ell n r = \frac{(1-\nu_1)F}{\pi G_1} \\ t_1 &= \frac{N(\frac{1}{2}-\nu_1)}{\pi G_1} \left( \theta - \frac{\pi}{2} \right) = \frac{(\frac{1}{2}-\nu_1)F'}{\pi G_1} \end{aligned} \quad (35)$$

where  $F$  and  $F'$  are as in (34).

Similarly if  $p(x)=0$  and  $q(x)$  reduces to a concentrated shear force  $T$  at the origin, then eqs. (31) and (32) reduce to (fig. 7)

$$\begin{aligned} f &= 0, & g &= \ell n(Z/i) \\ G &= T \ell n r, & G' &= T \left( \theta - \frac{\pi}{2} \right) \end{aligned} \quad (36)$$

It can now be shown that, in terms of these harmonics, the boundary displacements are given by

$$n_1 = -\frac{(\frac{1}{2}-\nu_1)G'}{\pi G_1}, \quad t_1 = -\frac{(1-\nu_1)G}{\pi G_1} \quad (37)$$

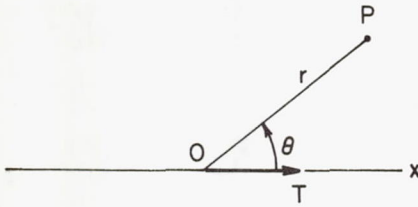


FIGURE 6.—Illustration for equations (34).

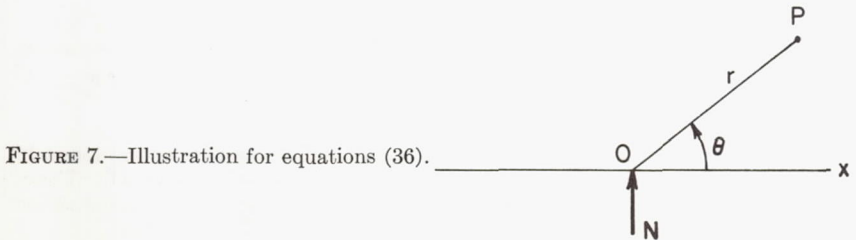


FIGURE 7.—Illustration for equations (36).

Applying superposition in passing from  $N$  to  $p(x)$  and from  $T$  to  $q(x)$  and adding, there results for the boundary displacements

$$\begin{aligned} n_1 &= -\frac{(1-\nu_1)F}{\pi G_1} - \frac{(\frac{1}{2}-\nu_1)G'}{\pi G_1} \\ t_1 &= \frac{(\frac{1}{2}-\nu_1)F'}{\pi G_1} - \frac{(1-\nu_1)G}{\pi G_1} \end{aligned} \quad (38)$$

where  $F$ ,  $F'$ ,  $G$ , and  $G'$  are given by eq. (32).

As stated, these displacement equations apply only on the boundary  $y=0$  and do not represent the displacements elsewhere.\* Nevertheless, it is of interest to consider the functions (32) for the whole half-plane  $y \geq 0$ . From eq. (32), it will be seen that  $F$  may be identified with the logarithmic potential of a charge distribution over the strip, (30), of density proportional to  $p(x)$ ; similarly  $F'$  corresponds to the flux function of the same field.

Since  $\ell n r$  in (32) is even in  $y$ , it follows that for  $y=0$ ,  $|x| > a$ ,  $F$  satisfies the relation

$$\frac{\partial F}{\partial y} = 0 \quad \text{for } y=0, |x| > a \quad (39)$$

while over the interval (30) it follows from properties of  $F$  as a potential that

$$\left. \frac{\partial F}{\partial y} \right|_{y=0^+} = \pi p(x) \quad \text{for } y=0^+, |x| < a \quad (40)$$

where  $y=0^+$  indicates the limit approached as  $y$  approaches 0 through positive values. It now follows from the Cauchy-Riemann equations for  $F$  and  $F'$  that (39) and (40) lead to

$$\begin{aligned} \frac{\partial F'}{\partial x} &= 0 & \text{for } y=0^+, |x| > a \\ \frac{\partial F'}{\partial x} &= -\pi p(x) & \text{for } y=0^+, |x| < a \end{aligned} \quad (41)$$

Integrating eq. (41) along  $y=0$  and noting from figure 5 that  $\theta$  approaches zero over  $x > a$  as  $y \rightarrow 0^+$ , the following boundary values result for  $F'$

---

\* To obtain the displacements over the whole half-plane  $y=0$  of figures 6 and 7, it is necessary to add multiples of the functions  $\cos 2\theta$ ,  $\sin 2\theta$  to (35) and (37). (These are actually independent of  $r$ .) However, over the boundary  $y=0$ , these terms either vanish or reduce to a constant; hence, they have been omitted in (35)-(38).

$$F' = \begin{cases} -P \frac{\pi}{2} & \text{for } y=0^+, x > a \\ P \frac{\pi}{2} - \pi \int_{-a}^x p(s) ds & \text{for } y=0^+, |x| < a \\ P \frac{\pi}{2} & \text{for } y=0^+, x < -a \end{cases} \quad (42)$$

where

$$P = \int_{-a}^a p(s) ds \quad (43)$$

This also follows directly from (32)<sub>1</sub> by replacing  $\theta$  by its limiting values 0 and  $\pi$  as  $y \rightarrow 0^+$  for  $s < x$  and  $s > x$ , respectively.

The functions  $F$  and  $F'$  may also be considered in the lower half-plane. Then  $F$  turns out to be a single-valued function, analytic in the entire plane except along the interval (30), i.e., in the plane slit along the interval (30), and even in  $y$ :

$$F(x, -y) = F(x, y) \quad (44)$$

On the other hand,  $F'$  takes on one constant value along AC on figure 5 and a different constant value over A'C'. When it is continued across  $y=0$  over AC and over A'C', one is led to a multiple-valued function that increases by  $2\pi P$  for a path that encircles A'OA positively.

By expanding  $\ln(Z-s)$  in eq. (31) in powers of  $s$  and integrating term-wise, one obtains the following series for  $f$ :

$$f(Z) = F + iF' = P \ln \left( \frac{Z}{i} \right) - \frac{m_1}{Z} + \dots \quad (45)$$

where

$$m_1 = \int_{-a}^a s p(s) ds \quad (46)$$

is the first moment  $p(x)$ , while the following terms involve higher-order moments of  $p(x)$ . Hence, if  $P \neq 0$ , the asymptotic behavior of  $F$  and  $F'$  is described by

$$F \sim P \ln \sqrt{x^2 + y^2}, \quad F' \sim P \left( \theta - \frac{\pi}{2} \right) \quad (47)$$

If  $P$  vanishes, then the asymptotic behavior at infinity is described by

$$F + iF' \sim -\frac{m_1}{Z}, \quad F \sim -\frac{m_1 \cos \theta}{r}, \quad F' \sim \frac{m_1 \sin \theta}{r} \quad (48)$$



The functions  $G$  and  $G'$  possess properties similar to those of  $F$  and  $F'$ , provided  $p(x)$  is replaced by  $q(x)$ ,  $P$  by

$$Q = \int_{-a}^a q(s) ds \quad (49)$$

and  $m_1$  by the moment of  $q(x)$ . In symbolic form we have the "proportion"

$$(F, F') : p = (G, G') : q \quad (50)$$

where ":" is to be interpreted as "is to" and "=" as "as."

While definite expressions have been obtained for the boundary displacements above, it will be recalled that an arbitrary rigid displacement can be added to any solution of the equations of elasticity. In particular, constants may be added to  $F$ ,  $F'$ ,  $G$ , and  $G'$  in the various equations above. Thus  $\ln r$  may be replaced by  $\ln(r/a)$ , rendering the argument of  $\ln$  dimensionless. Again,  $(\theta - \pi/2)$  may be replaced by  $\theta$ , thus adding the constant  $P\pi/2$  to the boundary values of  $F'$  in eq. (42).

It will be recalled that the stress distribution corresponding to the loading in figure 6 is given by

$$p_{rr} = -\frac{2N \sin \theta}{\pi r}, \quad p_{\theta\theta} = p_{r\theta} = 0 \quad (51)$$

while for the loading in figure 7 it is

$$p_{rr} = -\frac{2T \cos \theta}{\pi r}, \quad p_{\theta\theta} = p_{r\theta} = 0 \quad (52)$$

These sinusoidal tractions are indicated schematically in figures 8 and 9.

For continuous distributions  $p(x)$  and  $q(x)$ , eqs. (51) and (52) apply at large distances, but with  $N$  and  $T$  replaced by  $P$  and  $Q$ . These stress

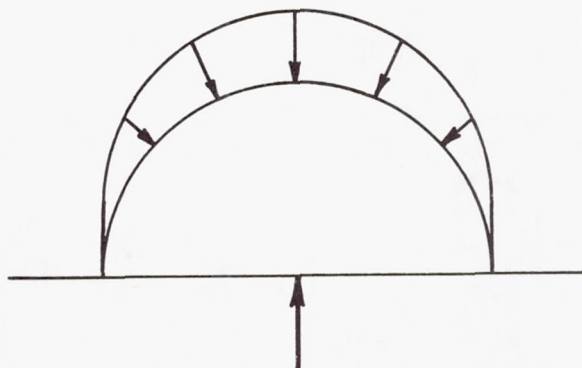


FIGURE 8.—Illustration of stresses, equation (51).

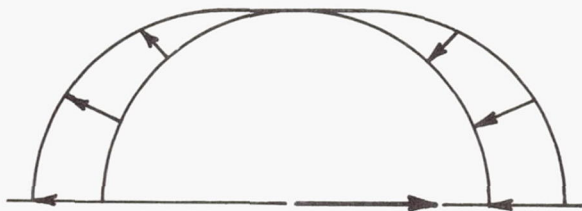


FIGURE 9.—Illustration of stresses, equation (52).

distributions are the only ones that vary inversely with  $r$ . They may or may not apply to a region extending to infinity; quite often they must be modified at appreciable distances from the Hertz contact zone. Thus, for a locomotive wheel (even assuming it to be a wheel of constant thickness), both  $P$  and  $Q$  are held in equilibrium by reactions applied by the axle. Again, for a gear tooth in mesh with a mating tooth, the applied (nearly normal) load is held in equilibrium by tractions at the base of the tooth while its tip side is idle. For two-dimensional contact problems this requires the introduction of more complex Airy stress functions.

We now return to the contact over the interval (30) of two cylindrical bodies shown schematically in figure 10 of radii  $R_1, R_2$ ; elastic constants  $G_1, \nu_1; G_2, \nu_2$ , with distributed tractions  $p(x), q(x)$ , positive as in figure 5 when they represent the tractions of the lower body on the upper one and are directed in direction of positive  $x, y$  while the upper body reacts with tractions  $-p(x), -q(x)$  on the lower one. Then, with  $y=0^+$ , eq. (38) yields the boundary displacements for the upper body, while for the lower one, with  $n_2, t_2$  measured positive as indicated in figure 10, they are given by

$$\begin{aligned} n_2 &= -\frac{(1-\nu_2)F}{\pi G_2} + \frac{(\frac{1}{2}-\nu_2)G'}{\pi G_2} \\ t_2 &= \frac{(\frac{1}{2}-\nu_2)F'}{\pi G_2} + \frac{(1-\nu_2)G}{\pi G_2} \end{aligned} \quad (53)$$

where  $F, F'$  and  $G, G'$  are the limiting values as  $y \rightarrow 0^+$ , of the same functions defined for  $y \geq 0$  by eq. (32).

Conformal mapping is often useful in solving for  $F, F', \dots$  in some contact problems. This mapping is effected by letting

$$Z = Z(\zeta), \quad Z = x + iy \quad \zeta = \xi + i\eta \quad (54)$$

where  $Z(\zeta)$  is an analytic function of  $\zeta$ . Eq. (54) transforms the upper half  $Z$ -plane into a proper region  $R$  of the  $\zeta$ -plane in a one-to-one manner, so that to each point in  $R$  corresponds one point in  $y \geq 0$ , and conversely.

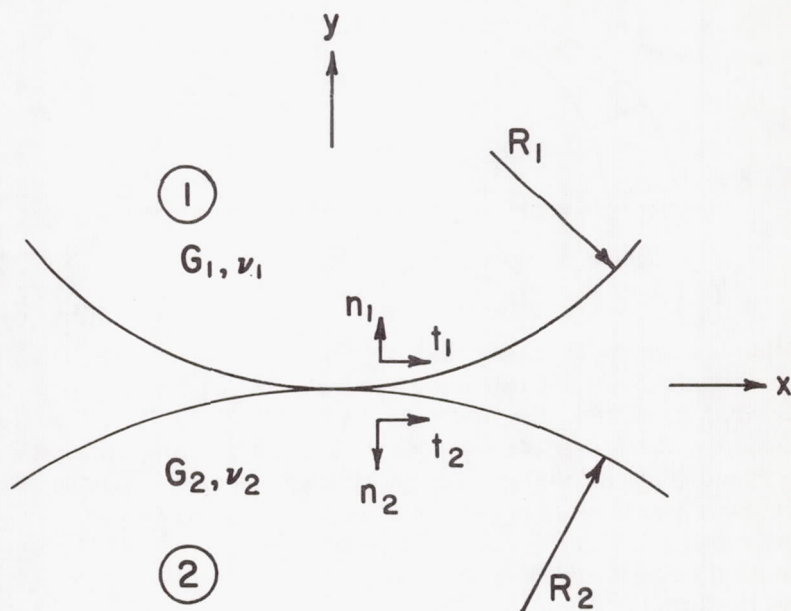


FIGURE 10.—Displacement sign conventions for contacting cylinders.

At the same time an analytic function of  $Z$  is transformed into an analytic function of  $\zeta$ . In particular the function  $f(Z) = F + iF'$  is transformed into an analytic function of  $\zeta$ , and  $F, F'$  into conjugate harmonic functions of  $\xi, \eta$ .

The term "conformal" is due to the property that angles are preserved. Indeed, taking differentials of (54), there results

$$dZ = \left( \frac{dZ}{d\zeta} \right) d\zeta \quad (55)$$

from which follows that, if  $dZ/d\zeta \neq 0$ , a small neighborhood near a point is magnified in the ratio  $dZ/d\zeta$  and rotated through an angle equal to the argument of  $dZ/d\zeta$ .

As an example, we consider the mapping effected by

$$Z = a \sin \zeta, \quad \zeta = \sin^{-1} \left( \frac{Z}{a} \right) = \int_0^Z \frac{dZ}{\sqrt{a^2 - Z^2}} \quad (56)$$

This maps the half plane  $y > 0$  of figure 5 conformally on the semi-infinite strip

$$-\frac{\pi}{2} < \xi < \frac{\pi}{2}, \quad \eta > 0 \quad (57)$$



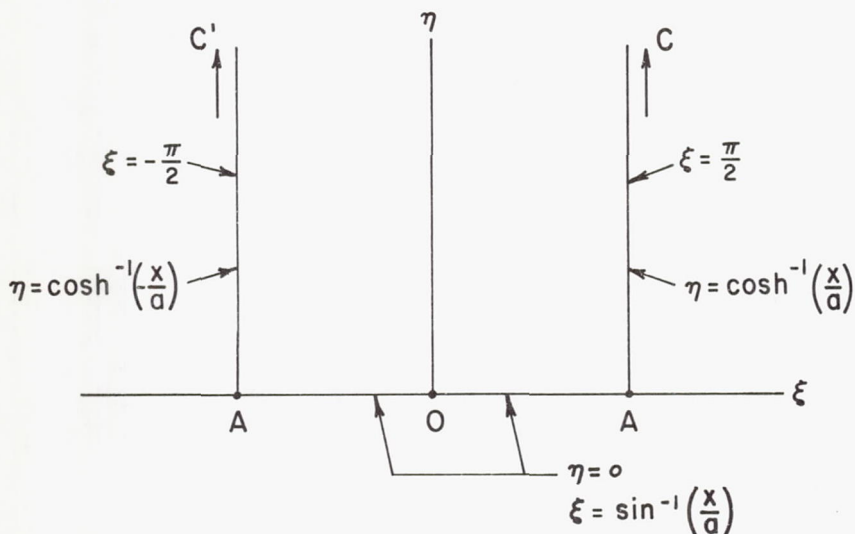


FIGURE 11.—Conformal mapping of half-plane  $y > 0$  on a semi-infinite strip  $-\pi/2 < \xi < \pi/2$ ,  $\eta = 0$ , showing boundary values of  $\xi$  and  $\eta$ .

shown in figure 11. Corresponding points in figures 5 and 11 have been similarly labelled. The conformality property is violated at the points  $A'$ ,  $A$ , where the rectilinear boundary  $y=0$  is "bent" into a right angle corner. These are precisely the points at which the derivative

$$\frac{dZ}{d\zeta} = a \cos \zeta \quad (58)$$

vanishes.

Expanding  $\sin(\xi + i\eta)$ , replacing  $\sin i\eta$ ,  $\cos i\eta$  by  $i \sinh \eta$ ,  $\cosh \eta$ , respectively, there results from eq. (56)

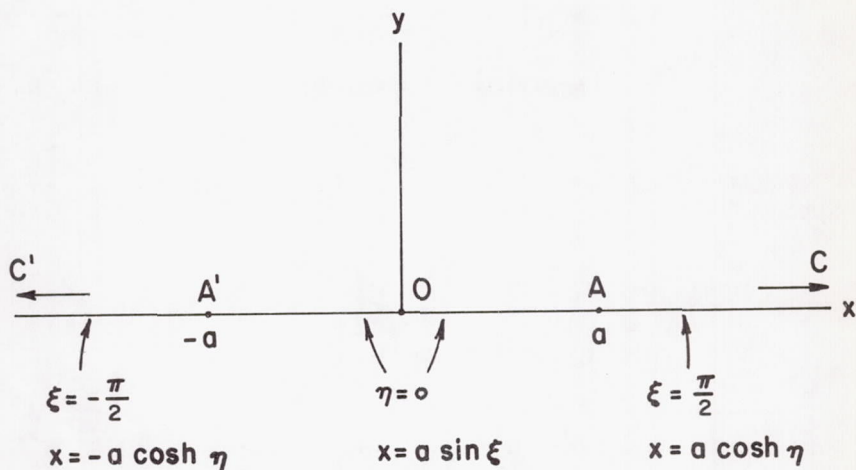
$$x = a \sin \xi \cosh \eta, \quad y = a \cos \xi \sinh \eta \quad (59)$$

Putting  $\xi$ ,  $\eta$  values corresponding to each side of the semi-infinite strip (57) yields  $y=0$  and values of  $x$  indicated in figure 12. Solving for  $\xi$  or  $\eta$  results in the relations indicated in figure 11.

We seek harmonic functions  $F$  that satisfy the condition (39) at the "free" boundary on  $|x| > a$ , and behave at infinity as indicated in (47). In the  $\zeta$ -plane, this leads to

$$\frac{\partial F}{\partial \xi} = 0 \quad \text{for} \quad \xi = -\frac{\pi}{2}, \frac{\pi}{2} \quad (60)$$

Harmonic functions of the product type in  $\xi$ ,  $\eta$  satisfying (60) are given by

FIGURE 12.—Boundary values of  $\xi$  and  $\eta$  in the  $(x, y)$ -plane.

$$F: \eta, \sin \xi e^{-\eta}, \cos 2\xi e^{-2\eta}, \sin 3\xi e^{-3\eta}, \dots \quad (61)$$

Of these the first one has the asymptotic behavior (47) (if multiplied by a proper constant); the other solutions vanish at infinity. The use of positive exponentials in functions (61) would lead to prohibitively large numerical behaviors at infinity.

The conjugate harmonics of the functions (61) are given by

$$F': -\xi, -\cos \xi e^{-\eta}, \sin 2\xi e^{-2\eta}, -\cos 3\xi e^{-3\eta}, \dots \quad (62)$$

and, indeed, these satisfy the conditions (41)<sub>1</sub>, which lead to constant  $F'$  over  $y=0$ ,  $|x| > a$ .

To illustrate the use of these solutions, consider a multiple of the first one:

$$F = P\eta, \quad F' = -P\xi, \quad f = F + iF' = -iP\xi \quad (63)$$

Over the boundary  $y=0$  of figure 12, there results

$$f = F + iF' = -iP(\xi + i\eta) = -iP \sin^{-1} \left( \frac{x}{a} \right) = -iP \int_0^x \frac{dx}{\sqrt{a^2 - x^2}} \quad (64)$$

and, utilizing the boundary relations of figures 11 and 12, we obtain the following boundary values of  $F$  and  $F'$  over  $y=0$

$$F = P_\eta = \begin{cases} P \cosh^{-1} \frac{|x|}{a} = P \ln \left( \frac{|x|}{a} + \sqrt{x^2 - a^2} \right), & x > a \\ 0, & |x| < a \end{cases} \quad (65)$$

$$F' = -P\xi = \begin{cases} P\pi/2 & \text{for } x < -a \\ -P \sin^{-1}(x/a) & \text{for } -a < x < a \\ -P\pi/2 & \text{for } x > a \end{cases} \quad (66)$$

From eqs. (41) and (42), it follows that this solution corresponds to an applied load  $P$ , distributed over  $|x| < a$  as follows:

$$p(x) = -\frac{1}{\pi} \frac{\partial F'}{\partial x} = \frac{P}{\pi} \frac{1}{\sqrt{a^2 - x^2}} \quad (67)$$

If we suppose, further, that there are no shearing stresses:

$$q(x) = 0, \quad G = 0, \quad G' = 0 \quad (68)$$

then eqs. (38) (upon the addition of a constant) yield

$$n_1 = -\frac{P(1-\nu_1)}{\pi G_1} \begin{cases} n_0 & \text{for } |x| < a \\ n_0 - \cosh^{-1} \frac{|x|}{a} & \text{for } |x| > a \end{cases} \quad (69)$$

Thus the normal displacement over the interval (30) is constant (see fig. 13).

Physically the solutions (63) through (69) correspond to a horizontal, flat, rigid die, of width  $2a$ , pressed against  $y=0$  over the interval (30) with a force  $P$  (per unit  $z$ ). Eq. (38) shows that the solution carries with it the tangential displacement over  $|x| < a$

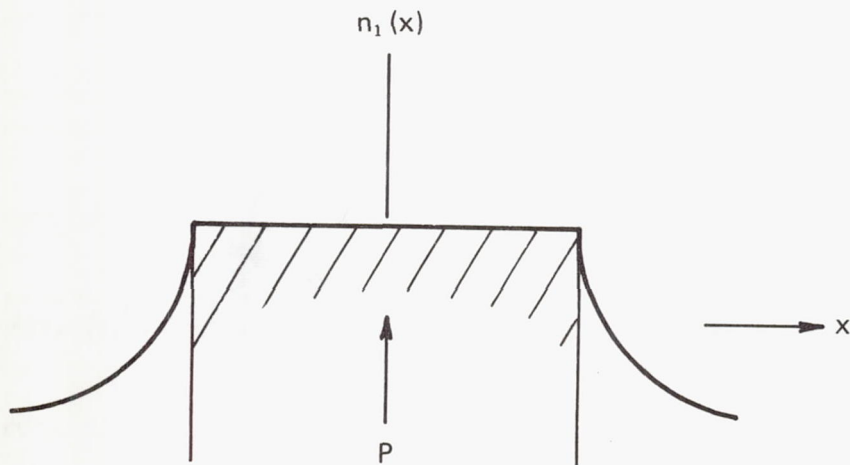


FIGURE 13.—Boundary distortion produced by a rigid die (schematic).



$$t_1 = \frac{(\frac{1}{2} - \nu_1)F'}{\pi G_1} = -\frac{P(\frac{1}{2} - \nu_1)}{\pi G_1} \sin^{-1} \frac{x}{a} \quad (70)$$

without arousing any shearing forces—presumably implying an ideal lubricant over the surface of the die.

As a second example, suppose that the above die is also rotated through an angle  $\alpha$  by the application of a moment  $M$ . It turns out that the solution to be added is a multiple of the second solution in the functions (62) and (63):

$$f = F + iF' = 2Mie^{i\xi} = 2Mie^{(i\xi - \eta)} \quad (71)$$

whence

$$F = -2M \sin \xi e^{-\eta}, \quad F' = 2M \cos \xi e^{-\eta} \quad (72)$$

Again, assuming conditions (68), there results for  $y=0$

$$n_1 = \frac{2M(1 - \nu_1)}{\pi G_1 a} \begin{cases} a \sin \xi = x & \text{for } |x| < a \\ ae^{-\eta} = x - \sqrt{x^2 - a^2} & \text{for } x > a \\ -ae^{-\eta} = -x + \sqrt{x^2 - a^2} & \text{for } x < -a \end{cases} \quad (73)$$

Over  $|x| < a$  this corresponds to a rotation of the die through an angle

$$\alpha = \frac{2M(1 - \nu_1)}{\pi G_1 a} \quad (74)$$

The pressure distribution over the contact interval is found from eqs. (41) and (72):

$$p(x) = -\frac{2M}{\pi a} \frac{\partial \sqrt{a^2 - x^2}}{\partial x} = \frac{2M}{\pi a} \frac{x}{\sqrt{a^2 - x^2}} \quad (75)$$

and has a zero resultant and a moment  $m_1 = M$  about  $x=0$ , corresponding to the asymptotic behavior (48). The solution also implies a tangential displacement

$$t_1 = \frac{(1 - 2\nu_1)M}{\pi G_1 a} \sqrt{a^2 - x^2} \quad \text{for } |x| < a \quad (76)$$

It will now be shown that the Hertz solution is obtained by choosing for  $F, F'$  a proper combination of the first and third solutions (61), (62):

$$\begin{aligned} F &= P(\eta - \cos 2\xi e^{-2\eta}/2) \\ F' &= -P(\xi + \sin 2\xi e^{-2\eta}/2) \end{aligned} \quad (77)$$

Indeed, the pressure distribution, eq. (2), combined with eqs. (41) and (42), leads over interval (30) to

$$F' \big|_{v=0} = -\pi \int_0^x p(x) dx = -\frac{2P}{a^2} \int_0^x \sqrt{a^2 - x^2} dx \quad (78)$$

and, replacing  $x$  by  $a \sin \xi$  and integrating, to

$$F' = -2P \int_0^\xi \cos^2 \xi d\xi = -P \left( \xi + \frac{\sin 2\xi}{2} \right) \quad (79)$$

which agrees with  $(77)_2$  for  $\eta=0$ , and since  $F'$  in  $(77)_2$  also satisfies the conditions  $(42)_1$ ,  $(42)_3$ ,  $F'$  is determined as in  $(77)$  and so is its conjugate harmonic  $F$ .

The displacements  $n_1$ ,  $n_2$  over interval  $(30)$  are now found from eqs.  $(38)$  and  $(53)$  under the assumption  $(68)$ . First we note that along the  $x$ -axis and the strip boundary,  $F$  and  $F'$  in  $(77)$  take on the following boundary values:

$$F = P \left\{ \begin{array}{ll} -(\cos 2\xi)/2 & \text{for } \eta=0, \\ \eta + e^{-2\eta}/2 & \text{for } \xi = \pm \frac{\pi}{2} \end{array} \right\} = -PR_1 \left( \frac{x}{a} \right) \quad (80)$$

$$F' = -P \left\{ \begin{array}{ll} \xi + (\sin 2\xi)/2 & \text{for } \eta=0, \\ \pm \frac{\pi}{2} & \text{for } \xi = \pm \pi/2 \end{array} \right\} = -PR_2 \left( \frac{x}{a} \right)$$

where, substituting from figures 11 and 12 and expressing in terms of  $x$ ,

$$F = P \left\{ \begin{array}{ll} -\frac{1}{2} + \frac{x^2}{a^2} & \text{for } |x| < a \\ \ln \left( \frac{|x|}{a} + \sqrt{\frac{x^2}{a^2} - 1} \right) + \frac{1}{2} \left( \frac{x}{a} - \sqrt{\frac{x^2}{a^2} - 1} \right)^2 & \text{for } |x| > a \end{array} \right. \quad (81)$$

$$F' = -P \left\{ \begin{array}{ll} \sin^{-1} \frac{x}{a} + x \frac{\sqrt{a^2 - x^2}}{a^2} & \text{for } |x| < a \\ \pm \frac{\pi}{2} & \text{for } \begin{cases} x > a \\ x < -a \end{cases} \end{array} \right.$$

Now apply eqs.  $(38)$  and  $(53)$  and replace  $G$  and  $G'$  by zero. There results

$$\begin{aligned}
 n_1 &= \frac{P(1-\nu_1)}{\pi G_1} R_1 \left( \frac{x}{a} \right), & t_1 &= -\frac{P(\frac{1}{2}-\nu_1)}{\pi G_1} R_2 \left( \frac{x}{a} \right) \\
 n_2 &= \frac{P(1-\nu_2)}{\pi G_2} R_1 \left( \frac{x}{a} \right), & t_2 &= -\frac{P(\frac{1}{2}-\nu_2)}{\pi G_2} R_2 \left( \frac{x}{a} \right)
 \end{aligned}
 \tag{82}$$

Eq. (1) is now obtained by applying the condition of avoiding interpenetration over  $|x| < a$

$$n_1 + n_2 = -\frac{x^2}{2} \left( \frac{1}{R_1} + \frac{1}{R_2} \right) + \text{const} \tag{83}$$

Comparison of  $t_1$  and  $t_2$  in eqs. (38) and (53) shows that only if the solids are similar elastically

$$G_1 = G_2, \quad \nu_1 = \nu_2 \tag{84}$$

is there no slip between the solids. Otherwise the solution implies slip of total magnitude

$$t_1 - t_2 = \frac{P}{\pi} \left( \frac{\frac{1}{2}-\nu_1}{G_1} - \frac{\frac{1}{2}-\nu_2}{G_2} \right) R_2 \left( \frac{x}{a} \right) \tag{85}$$

over  $|x| < a$ . In general, these will give rise to shearing frictional forces that have been neglected so far.

Summarizing, the ordinary Hertz theory is valid if the solids are similar elastically; otherwise slip must occur over the contact area.

When shearing tractions  $q(x)$  do occur as a result of the above effects produced by the pressure, and due to applied tangential tractions, then one must add to the above the displacements due to  $G$  and  $G'$  in eqs. (38) and (53):

$$\begin{aligned}
 n_1 &= -\frac{(\frac{1}{2}-\nu_1)G'}{\pi G_1}, & t_1 &= -\frac{(1-\nu_1)G}{\pi G_1} \\
 n_2 &= \frac{(\frac{1}{2}-\nu_2)G'}{\pi G_2}, & t_2 &= \frac{(1-\nu_2)G}{\pi G_2}
 \end{aligned}
 \tag{86}$$

It will be noted from the above that for similar solids, the contact area will be warped, but to the same shape (since  $n_1$  and  $n_2$  are measured positive in opposite directions; see fig. 10). For dissimilar solids, eq. (86) implies a gap between the solids over the contact area, which generally requires a redistribution of the pressure to close it. Whether the solids are similar or not, eq. (86) implies added slip, which may cancel some of that given by eq. (85).



We return to the assumptions (68) and (84). As explained in reference 2, for an arbitrary pressure distribution, one may expand  $F'$  in a series of trigonometric functions displayed in (62) (along with the  $\xi$ -term). Then the displacements can be found from the corresponding trigonometric series from the function (61). Conversely, given the displacement  $n_1$  over interval (30),  $F$  is obtained from the trigonometric series derived from (61) and the pressure distribution for a given load  $P$  obtained from the series (62) with the same coefficients (along with the  $\xi$ -term as in eq. (66)). If we introduce

$$\xi' = \xi - \frac{\pi}{2} \quad (87)$$

then the solutions (61) and (62) become

$$\begin{aligned} F: \quad & \eta, \quad \cos \xi' e^{-\eta}, \quad \cos 2\xi' e^{-2\eta}, \quad \cos 3\xi' e^{-3\eta}, \dots \\ F: \quad & -\xi', \quad \sin \xi' e^{-\eta}, \quad \sin 2\xi' e^{-2\eta}, \quad \sin 3\xi' e^{-3\eta}, \dots \end{aligned} \quad (88)$$

so that the Fourier expansions of  $F$  and  $F'$  over the interval  $0 < \xi' < \pi$  reduce to Fourier cosine series and Fourier sine series, respectively.

The sine functions in (61) can be written as polynomials in  $\sin \xi$  and hence as polynomials in  $x$ , which turn out to be the Tschebycheff polynomials  $T_1, T_3, \dots$  in  $x/a$ . Likewise, the cosine terms of (61) may be put in the form  $\cos \xi \times (\text{a polynomial in } \sin \xi)$ ; hence they can be expressed as  $[1 - (x/a)^2]^{1/2} \times (\text{polynomials in } x/a)$ . The advantages of the conformal mapping with its Fourier series is the greater ease of the Fourier series expansions and the calculation of the displacements outside the contact zone.

Alternative conformal transformations are of interest. The mapping

$$Z = \frac{a}{2} \left( \zeta_1 + \frac{1}{\zeta_1} \right) \quad (89)$$

where

$$\zeta_1 = \rho_1 e^{i\phi_1} \quad (90)$$

maps the  $Z$ -plane, slit along the interval (30), into the exterior of the unit circle  $\rho_1 = 1$  in the  $\zeta_1$ -plane, and the upper half  $Z$ -plane into the portion of the upper-half  $\zeta_1$ -plane exterior to this circle. The functions  $f$  corresponding to the solutions (61), (62) now reduce to

$$\ell n \zeta_1, \quad \zeta_1^{-1}, \quad \zeta_1^{-2}, \dots \quad (91)$$

Other conformal transformations are useful in solving multiple-contact problems. For two flat dies of the same size, pressed against  $y=0$  over  $b < |x| < c$  so as to produce the same constant deflection, the transformation

$$Z_1 = Z^2 \quad (92)$$

mapping the first quadrant in the  $Z$ -plane on the upper-half of the  $Z_1$ -plane, essentially reduces the problem to one of a single die, and leads to the pressure distribution

$$p(x) = \frac{2P |x|}{\sqrt{(x^2 - b^2)(c^2 - x^2)}} \quad \text{over } b < |x| < c \quad (93)$$

For more general double-contact problems, where  $F'$  is constant between the contact areas and  $F$  is prescribed over the contacts, conformal mapping of the upper half  $Z$ -plane on a rectangle by means of elliptic integrals is useful in rendering the solutions more tractable. In some cases, the solution itself can be expressed as an elliptic integral in  $Z$  or in a variable related to it by means of a linear fractional transformation. In other cases, series of product solutions in the plane of the rectangle can be used to solve for  $F$  and  $F'$ . Some of the above-mentioned integrals are not unlike those occurring in stress concentration solutions near a crack.

### 3. CONTACTS OF ELASTICALLY SIMILAR CYLINDERS WITH NORMAL AND TANGENTIAL TRACTIONS

First we consider the case of tractive rolling of a locomotive driving wheel, or of two cylinders interacting with a normal force  $P$  and a tangential force  $Q$  satisfying the inequality (3)—a problem for which the solution was outlined in the first section.

As indicated in the second section, the effect of the shearing stress  $q(x)$  and the functions  $G, G'$  in eqs. (38) and (53) leads to a warping of the area of contact in addition to its flattening as expressed in eq. (83). Under the assumptions (84), the warping is the same for both the track and the wheel and does not require any modifications in eqs. (2) and (77) through (81).

To examine the slip and creep, we differentiate with respect to  $x$  the eqs. (38)<sub>2</sub>, (53)<sub>2</sub> after first substituting from eq. (32) for  $G$  and utilize eq. (41). There results

$$(e_x)_1 = \frac{\partial t_1}{\partial x} = -\frac{(\frac{1}{2} - \nu_1)}{G_1} p(x) - \frac{(1 - \nu_1)}{\pi G_1} \int_{-a}^a \frac{q(s) ds}{x - s} = (e_{xp})_1 + (e_{xq})_1 \quad (94)$$

$$(e_x)_2 = \frac{\partial t_2}{\partial x} = -\frac{(\frac{1}{2} - \nu_2)}{G_2} p(x) + \frac{(1 - \nu_2)}{\pi G_2} \int_{-a}^a \frac{q(s) ds}{x - s} = (e_{xp})_2 + (e_{xq})_2$$

where  $(e_{xp})$ 's represent the components of strain produced by pressure (and which vanish outside the contact interval) and  $(e_{xq})$ 's, those by a shear. When the contacting bodies are similar elastically, the  $e_{xp}$  com-

ponents for both of them are equal, while the  $e_{xq}$  components are equal and opposite. This verifies the statements regarding eqs. (13) and (24) made in the first section.

Over the locked area  $L$ , as indicated in figure 3, each  $e_{xq}$  must be constant. Utilizing the symbolic analogy (50), it will be seen that the component of  $q(x)$  due to the first term in eq. (49) and represented by the outer circle in figure 4 gives rise to  $G, G'$  functions obtainable from (77) by multiplication by  $\mu$ . The component due to the last term in eq. (23) and represented by the smaller semicircle in figure 4 can be obtained by shifting the center of the circle to  $x = a - b$ , replacing the radius  $a$  by  $b$ , and  $P$  by  $-Pb^2/a^2$ . Hence there results

$$G = \mu \left[ F(\xi, \eta) - \frac{b^2}{a^2} F(\xi_1, \eta_1) \right] \quad (95)$$

$$G' = \mu \left[ F'(\xi, \eta) - \frac{b^2}{a^2} F'(\xi_1, \eta_1) \right]$$

where

$$Z - (a - b) = b \sin \zeta_1, \quad \zeta_1 = \xi_1 + i\eta_1 \quad (96)$$

and where  $F(\xi_1, \eta_1), F'(\xi_1, \eta_1)$  are obtained by carrying out the replacements

$$\begin{aligned} \xi &\rightarrow \xi_1, & \eta &\rightarrow \eta_1 \\ x &\rightarrow x - a + b = x_1, & a &\rightarrow b, & y &\rightarrow y_1 \end{aligned} \quad (97)$$

implying the conformal mapping of the upper-half  $Z$ -plane on the semi-infinite strip,  $\eta_1 = 0, -\pi/2 < \xi_1 < \pi/2$ , with the diameter of the smaller circle going into the finite side.

It is now possible to resort to eqs. (80), (81) with corresponding equations for  $F(\xi_1, \eta_1), F'(\xi_1, \eta_1)$  to express eq. (95) in terms of  $x$ , and by differentiating with respect to  $x$ , obtain the values of  $(e_{xq})_1$ . This turns out to be relatively simple in the region  $L$  where eq. (81) indicates that  $t_1$  is quadratic in  $x$ , and  $(e_{xq})_1$  is linear in  $x$ ; it turns out that the replacement (97) also lead to a linear term with the same coefficient of  $x$ , and hence to a constant  $(e_{xq})_1$ . In  $S$  it is simpler to handle the  $F(\xi_1, \eta_1)$ -term by using the form derived from (77), the applying, for  $\xi_1 = -\pi/2$ ,

$$\begin{aligned} \frac{\partial F(\xi_1, \eta_1)}{\partial x} &= \frac{\partial F}{\partial \eta} \bigg/ \frac{\partial x}{\partial \eta} = \frac{P(1 - e^{-2\eta_1})}{-b \sinh \eta_1} \\ &= -\frac{2P}{b} e^{-\eta_1} = -\frac{2P}{b} \left( \frac{x_1}{b} - \sqrt{\left( \frac{x_1}{b} \right)^2 - 1} \right) \end{aligned} \quad (98)$$

After carrying out the proper substitutions into (95) and the pertinent



equations of the second section, we arrive over the contact interval at the eqs. (24), (25).

It is of interest to note that in the treatment in reference 2, I was not aware of the component of strain due to the pressure. The infinite value of the strain at the forward end of  $S$  has been noticed by Buefner in his discussion of Johnson's paper (ref. 4). Since linear elasticity is based on the assumption of small strains and displacements, infinite values derived on its basis should be taken with a grain of salt.

As a further example, we consider a stationary contact of elastically similar cylinders, which, in addition to the normal load  $P$  with its Hertzian solution, is also transmitting a tangential load  $Q$ , and assume that the coefficient of friction is infinite so that the horizontal displacement due to  $q(x)$  vanishes over the contact interval. A little reflection, utilizing the analogy (50), shows that (63) to (67) may now be applied to yield  $G$ ,  $G'$  by merely replacing  $P$  by  $Q$ . In particular there results for the shear distribution

$$q(x) = \frac{Q}{\pi} \frac{1}{\sqrt{a^2 - x^2}} \quad (99)$$

This agrees with a solution by Mindlin (ref. 7) for two solids with a Hertzian pressure distribution over an elliptic area of contact, where the two bodies also interact with a tangential force  $S_x$ ; under the assumption of zero slip, Mindlin showed that the shear distribution is given by

$$q(x) = \frac{3S_x}{2\pi ab} \left( 1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} \right)^{-1/2} \quad (100)$$

over the Hertzian ellipse (see sixth section). The agreement is in the sense that if  $b$  is allowed to become infinite, while

$$\lim_{b \rightarrow \infty} \frac{3S_x}{2b} = -Q \quad (101)$$

then  $S_x$  in the shear distribution (100) will approach (99).

The solution corresponding to the stress distribution (99) may be added to the Hertzian solution, thus yielding the solution corresponding to the contact of two cylinders under normal load  $P$  and transverse load  $Q$ , but only if the coefficient of friction is infinite so as not to allow any slip over the contact interval. For a stationary (i.e., non-rolling) contact and a finite coefficient of friction, as pointed out by Johnson (ref. 4), the solution of this problem can be obtained by displacing the smaller circle of figure 4 leading to the shear stress distribution

$$q(x) = \frac{2P\mu}{a^2} \begin{cases} \sqrt{a^2 - x^2} & \text{for } b < |x| < a \\ \sqrt{a^2 - x^2} - \sqrt{b^2 - x^2} & \text{for } |x| < b \end{cases} \quad (102)$$

where

$$Q = \mu P \left( 1 - \frac{b^2}{a^2} \right) \quad (103)$$

As a third and last example, we consider the contact of cylinders whose profiles are of the form

$$y = \pm (k_1 x^2 + k_2 x^4 + \dots) \quad (104)$$

The parabolic term is, of course, what has been used previously as an approximation to the circular shape. The fourth-degree term might be visualized as a change of shape from the circular profile, but is also of interest as a better approximation to the circular profile when the width of the contact zone is large enough so that the parabolic approximation is not accurate enough. In the latter case

$$k_1 = \frac{1}{2R^2}, \quad k_2 = \frac{1}{8R^3}, \dots \quad (105)$$

It must be clearly understood that if  $a/R$  is not small, then the assumption of small displacements and slopes is violated, and the replacement of the boundary beyond the contact zone by a straight boundary is quite unjustified. Thus one would not be justified in squashing the cylinders with so much force that they flatten over a contact zone extending  $20^\circ$  to each side of the initial common normal.

Confining ourselves to the  $k_1$ -,  $k_2$ -terms in (104) and assuming that there are no shears, it follows from eq. (38) that

$$n_1 = n_0 - k_1 x^2 - k_2 x^4 = -\frac{(1 - \nu_1)}{\pi G_1} F \quad \text{for } |x| < a \quad (106)$$

To find  $F'$  and the resulting pressure, we recall the conformal mapping (56) and figure 12, whence

$$n_1 = n_0 - k_1 a^2 \sin^2 \xi - k_2 x^4 \sin^4 \xi \quad (107)$$

and, replacing  $\sin \xi$  by  $(e^{i\xi} - e^{-i\xi})/2i$ , introducing terms  $\cos 2\xi$ ,  $\cos 4\xi$ , and dropping the additive constants,

$$n_1 = K_1 \cos 2\xi + K_2 \cos 4\xi \quad (108)$$

where

$$K_1 = \frac{k_1 a^2}{2} - \frac{3k_2 a^4}{8}, \quad K_2 = \frac{k_2 a^4}{8}$$

It now follows that  $F$  can be formed as a linear combination of the third and fifth solutions (61). Forming  $F'$  from the corresponding conjugate harmonics (62), along with the  $-P\xi$  term, there results

$$F = P\eta - \frac{\pi G_1}{1 - \nu_1} (K_1 \cos 2\xi e^{-2\eta} + K_2 \cos 4\xi e^{-4\eta}) \quad (109)$$

$$F' = -P\xi - \frac{\pi G_1}{1 - \nu_1} (K_1 \sin 2\xi e^{-2\eta} + K_2 \sin 4\xi e^{-4\eta})$$

The pressure  $p(x)$  is now obtained from eq. (41)

$$p(x) = -\frac{1}{\pi} \frac{\partial F'}{\partial x} = -\frac{1}{\pi} \frac{\partial F'}{\partial \xi} \bigg/ \frac{\partial x}{\partial \xi}$$

$$= \frac{P}{a\pi \cos \xi} + \left( \frac{G_1}{1 - \nu_1} \right) \left( \frac{2K_1 \cos 2\xi + 4K_2 \cos 4\xi}{a \cos \xi} \right) \quad (110)$$

Unless the numerator vanishes at  $x=a$ ,  $p(x)$  at the interval ends would become infinite as in (67). To obtain a Hertzian behavior we therefore put

$$-2K_1 + 4K_2 + \frac{P(1 - \nu_1)}{\pi G_1} = 0 \quad (111)$$

Substituting for  $K_1$ ,  $K_2$  from (108) yields

$$k_1 a^2 - \frac{5}{2} k_2 a^4 = \frac{P(1 - \nu_1)}{\pi G_1} \quad (112)$$

This is a quadratic in  $a^2$ , and its positive root yields the pertinent value of the size of the contact zone;  $p(x)$  is now determined from (110) and can be put in the form

$$p(x) = \sqrt{a^2 - x^2} (A_1 + A_2 x^2) \quad (113)$$

Similar corrections to the Hertz theory can be obtained for further powers of  $x^2$  in (104).

Even more intriguing is the inclusion of effects of  $q(x)$  and of the tangential displacements  $t_1$ ,  $t_2$ . The normal displacements due to  $q(x)$  produce the same warping of the contact interval; the tangential displacements due to  $p(x)$  and  $q(x)$ , lead, for stationary contacts, to a central locked zone  $L$  along which  $t(x)$  vanishes, and to outer slip zones where  $q(x) = \pm \mu p(x)$ .

Aside from the higher order corrections in the Hertz theory, these examples are of interest from the point of view of minimizing the slip losses, and form the two-dimensional analogue of the work by Mow, Chow, and Ling (ref. 8).

#### 4. CONTACT PHENOMENA OF ELASTICALLY DISSIMILAR CYLINDERS

We now consider contact phenomena of elastically dissimilar cylinders,



$G_1 \neq G_2$ ,  $\nu_1 \neq \nu_2$ , supposing that

$$\frac{1-2\nu_1}{G_1} > \frac{1-2\nu_2}{G_2} \quad (114)$$

Since the variation in  $\nu$  is usually not appreciable, eq. (114) implies that of the two rollers in figure 10, the upper one is the more resilient one.

As indicated in the second and third sections, for dissimilar cylinders the contact phenomena are more complex than for similar cylinders due to the fact that in eqs. (38), (53) the pressure  $p(x)$  affects the tangential displacements  $t_1$ ,  $t_2$  differently, and the shearing forces  $\pm q(x)$  warp the contact area differently in bodies 1 and 2. The  $(n_1+n_2)$ -condition for contact over  $|x| < a$  is obtained from eqs. (38), (53), (79), leading to

$$F + \alpha G' = -Dh(x+x_m) \quad (115)$$

where

$$\alpha = \left( \frac{\frac{1}{2} - \nu_1}{G_1} - \frac{\frac{1}{2} - \nu_2}{G_2} \right) / \left( \frac{1 - \nu_1}{G_1} + \frac{1 - \nu_2}{G_2} \right) = \frac{\kappa}{2} \quad (116)$$

$$D = -\pi / \left( \frac{1 - \nu_1}{G_1} + \frac{1 - \nu_2}{G_2} \right)$$

and where  $h(x)$ , given by the right-hand member of (83), has been shifted since the line of centers may not pass through the midpoint of the contact interval—an effect brought about by the action of  $q$  and the asymmetry of  $p(x)$ .\*

A further relation involving  $t_1$ ,  $t_2$  and  $p(x)$ ,  $q(x)$  takes different forms over  $L$ , the locked, and  $S$ , the slip portions of the contact zone. For steady rolling over  $L$ , the strain difference is constant

$$L: \quad e_{x1} - e_{x2} = \frac{\partial t_1}{\partial x} - \frac{\partial t_2}{\partial x} = \text{const.} = -\xi U \quad (117)$$

where  $\xi$  is the "creep" coefficient. Upon integration this leads to

$$L: \quad \alpha F' - G = D(\xi x + C) \quad (118)$$

over  $L$ , where the inequality

$$L: \quad |q(x)| < \mu p(x) \quad (119)$$

must also hold. Over  $S$  slip takes place, and

\* An effect, generally disregarded, arises from the possible slanting of the line of centers, leading to a relative shift of the two contacting profiles. In view of the quadratic approximation to these profiles implied in (83), such a shift is equivalent to a change in  $x_m$ .

$$S: q(x) = \pm \mu p(x) \quad (120)$$

depending on the direction of the slip

$$S: \frac{de_1}{dx} - \frac{de_2}{dx} \leq 0 \quad (121)$$

Eq. (120) may be replaced by

$$S: G' = \pm \mu F' + \text{const} \quad (122)$$

Analytic solutions of the above relations have been obtained only for several special cases, and these we proceed to review.

First we consider complete slip over  $|x| < a$ , when

$$Q = \mu P \quad (123)$$

Eq. (120) now reduces to

$$q(x) = \mu p(x) \quad \text{for } |x| < a \quad (124)$$

Hence

$$G = \mu F, \quad G' = \mu F' \quad (125)$$

and (115) reduces to

$$F + \mu \alpha F' = -Dh(x + x_m) \quad (126)$$

A solution of this problem was presented in reference 2, but upon careful reexamination it turns out to be incorrect. It was based on an expansion of  $F, F'$  in powers of  $\mu$ .

$$F = F_0 + \mu F_1 + \mu^2 F_2 + \dots, \quad F' = F_0' + \mu F_1' + \dots \quad (127)$$

where  $F_0, F_0'$  correspond to the Hertz solution (77). Equating coefficients of like powers of  $\mu$  on both sides of (126) there result the recurrence equations

$$F_1 = -\alpha F_0', \quad F_2 = -\alpha F_1', \dots \quad (128)$$

With  $F_1$  determined from (128)<sub>1</sub>,  $F_2$  is obtained from (128)<sub>2</sub> as

$$F_2 = \alpha F_1' \quad (129)$$

and, taking conjugates of (128)<sub>1</sub> and substituting in (129),

$$F_2 = \alpha^2 F_0'' \quad (130)$$

where  $F_0''$  is the "conjugate of the conjugate" of  $F_0$ . Now multiplication of  $F + iF'$  by  $i$  yields  $iF - F'$  of which  $-F'$  is the real part,  $iF$  the imaginary part; hence, the conjugate harmonic of the conjugate harmonic of a harmonic function  $F$  is  $-F$ , and eq. (130) yields

$$F_2 = -\alpha^2 F_0 \quad (131)$$

The procedure continued to the subsequent equation (128) leads to two geometric series which are readily summed.

The error in the above lies in "taking conjugates." It is quite true that the boundary values of a harmonic function over the boundary of a region  $R$  determine the function completely inside  $R$ , as well as its conjugate harmonic (the latter to within an arbitrary additive constant). However, eqs. (126), (127) hold not over the complete boundary  $y=0$  of the region  $y>0$ , but only over  $|x|<a$ . They do not hold over  $|x|>a$ ,  $y=0$ ; there the boundary conditions are readily seen to be

$$F_1' = F_2' = \dots = 0 \quad (132)$$

Hence one may not equate conjugate harmonics of eq. (129). The solution of  $F_2, F_3, \dots$  from (128) is more involved.

While the solution of reference 2 is incorrect, it is worth pointing out that an approximate solution of (126) by means of a finite number of terms in (127), chosen as multiples of the solutions (61), (62), and satisfying (126) by collocation or by equating various moments of both sides of (126), is possible.

Returning to eq. (126) we substitute for  $F, F'$  from (32), differentiate with respect to  $x$ , and recall eq. (41). There results the integral equation for  $p(x)$

$$\int_{-a}^a \frac{p(s)}{x-s} ds - \pi\mu\alpha p(x) = B(x+x_m) \quad (133)$$

$$B = \left( \frac{1}{R_1} + \frac{1}{R_2} \right) / \left( \frac{1-\nu_1}{G_1} + \frac{1-\nu_2}{G_2} \right)$$

where the integral is to be understood as the Cauchy principal value, i.e., as

$$\lim_{\epsilon \rightarrow 0} \left( \int_{-a}^{x-\epsilon} + \int_{x+\epsilon}^a \right)$$

[Similar remarks apply to the integrals in (94).]

A finite-form solution of (47) has been obtained by Desoyer (ref. 9) and Bufler (ref. 10), and is given by

$$p(x) = \frac{B}{\pi \sqrt{1 + (\alpha\mu)^2}} (a+x)^{1-\theta} (a-x)^{\theta} \quad (134)$$

where

$$\tan \pi\theta = -\frac{1}{\mu\alpha}, \quad \frac{1}{2} < \theta < 1$$

$$a^2 = \frac{P}{2\theta(1-\theta)B} \quad (135)$$

$$x_m = a(2\theta - 1)$$



As a further example of (126), we consider the case of a constant  $h$ ,  $h = h_0$ . This corresponds to the loading of the solid  $y > 0$  by means of a rigid, horizontal die, exerting a normal force  $P$  and a maximum shearing force  $Q$  satisfying (123) so that complete slipping results.

We consider the  $f = (F + iF')$ -plane. In this plane eq. (126), which holds over the contact interval  $A'O A$  on figure 12, yields a straight line of slope

$$\frac{dF'}{dF} = -\frac{1}{\mu\alpha} = -\tan \delta, \quad \delta = \pi(1 - \Theta) \quad (136)$$

On the other hand, from eqs. (42)<sub>1</sub>, (42)<sub>3</sub>, it follows that the images of the  $C'A'$ ,  $CA$  of figure 12, in the  $f$ -plane, are the horizontal lines  $F' = \frac{P\pi}{2}$ ,

$F' = -\frac{P\pi}{2}$  (see fig. 14). Since  $F$ ,  $F'$  are conjugate harmonics in  $y = 0$ ,

$f$  is an analytic function of  $Z$ , and the mapping of the  $Z$ -half plane on the  $f$ -plane is conformal, with the region  $y > 0$  going into the interior of the semi-infinite strip of figure 14 with the slanting base. Conformal mapping of a half-plane on the interior of a polygon is carried out by means of a "Schwartz-Christoffel" integral which is the case given by

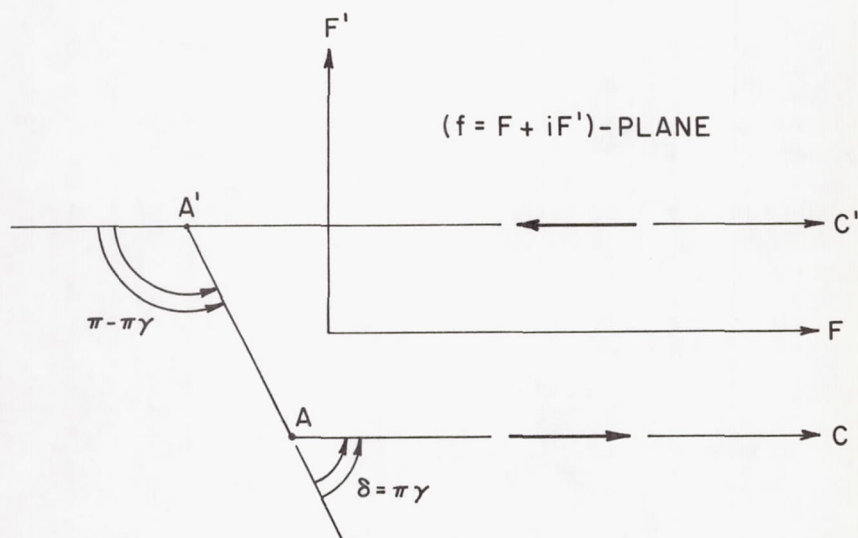


FIGURE 14.—Conformal mapping of the half-plane  $y > 0$  on the  $(F, F')$ -plane effected by equation (137).

$$f = F + iF' = A + P \int^Z (Z+a)^{\gamma-1} (Z-a)^{-\gamma} dZ \quad (137)$$

where  $Z = -a$ ,  $a$  corresponds to the vertices  $A'$ ,  $A$  on figure 14, and  $\pi\gamma$ ,  $\pi - \pi\gamma$  are the exterior angles at the vertices (indicated by double arcs). [Another instance of such mapping, for  $\gamma = \frac{1}{2}$  is that implied in figures 11 and 12, and eq. (56)<sub>2</sub>.]

In (137) the multiple-valued integrand is chosen so that for  $x > a$  along the  $x$ -axis it is positive, and so that the arguments of  $Z - a$ ,  $Z + a$  change from 0 to  $\pi$  as  $Z$  describes the upper plane while avoiding the neighborhoods of  $a$ ,  $-a$  by excluding them by means of small circles, as shown in figure 15. The arrows on the sides of the polygon of figure 14 correspond to description of the  $x$ -axis on figure 15 to the right.

Along the boundary  $y=0$ , there results

$$f(x) = F + iF'|_{y=0} = A + P \begin{cases} \int^x (x+a)^{\gamma-1} (x-a)^{-\gamma} dx, & x > a \\ e^{-i\pi\gamma} \int^x (x+a)^{\gamma-1} (a-x)^{-\gamma} dx, & x < a \\ - \int^x (-a-x)^{\gamma-1} (a-x)^{-\gamma} dx, & x < -a \end{cases} \quad (138)$$

where all the integrals are now positive.

The asymptotic behavior (47) for  $F$  is readily verified for (137) since the integrand reduces to  $Z^{-1}$  for large  $|Z|$ , also for  $F'$  if the constant is omitted and the indefinite integral in (137) is replaced by the definite

integral  $\int_{\bar{x}}^x$  where  $\bar{x}$  satisfies the relation

$$\int_{-\bar{a}}^{\bar{x}} (x+a)^{\gamma-1} (a-x)^{-\gamma} dx = \int_{\bar{x}}^a (x+a)^{\gamma-1} (a-x)^{-\gamma} dx \quad (139)$$

We now obtain for  $|x| < a$ ,  $y=0$

$$F = P \cos \pi\gamma \int_{\bar{x}}^x (x+a)^{\gamma-1} (a-x)^{-\gamma} dx \quad (140)$$

$$F' = -P \sin \pi\gamma \int_{\bar{x}}^x (x+a)^{\gamma-1} (a-x)^{-\gamma} dx$$

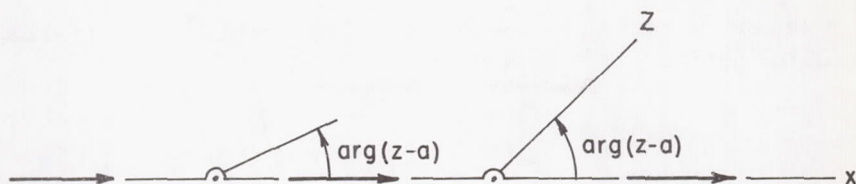


FIGURE 15.—Illustrating proper determination of multiple-valued integrand of equation (137).

and differentiating  $F'$  with respect to  $x$ ,

$$p(x) = \frac{P}{\pi} \sin \pi \gamma (x+a)^{\gamma-1} (a-x)^{-\gamma}, \quad q(x) = \mu p(x) \quad (141)$$

The shape of  $p(x)$  is shown schematically in figure 16.

Butler (ref. 10) also solves analytically the two integral equations obtained by differentiating the equations resulting from (115), (118), and (32) with respect to  $x$ , for the case where  $\mu$  is infinite and no slip, no creep can occur over the contact zone. By multiplying the second equation by  $i = \sqrt{-1}$ , adding to the first one, he converts the two integral equations into a single one for the complex-valued function

$$h(x) = p(x) + iq(x) \quad (142)$$

namely

$$h(x) - \frac{2i}{\pi \alpha} \int_{-a}^a \frac{h(s)}{x-s} ds = \ell(x) \quad (143)$$

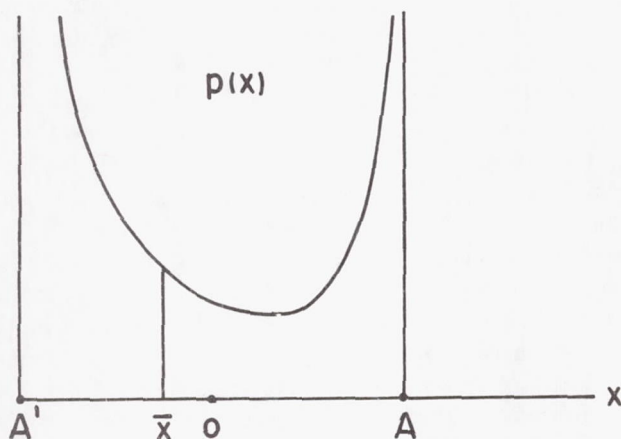


FIGURE 16.—Schematic pressure distribution, equation (141).



where the right-hand side is linear in  $x$ . He solves this by a method due to Söhngen (ref. 11), obtaining

$$p(x) + iq(x) = \frac{4P}{\pi a^2} \frac{\sqrt{a^2 - x^2}}{(1 + 4\beta^2) \sqrt{4 - \kappa^2}} \left( \frac{a+x}{a-x} \right)^{i\beta} \quad (144)$$

where

$$\beta = \frac{1}{2\pi} \ell n \left( \frac{2+\kappa}{2-\kappa} \right) \quad (145)$$

and  $a$  is given by

$$\frac{a}{a_h} = \frac{1}{\sqrt{1 + 4\beta^2}} \quad (146)$$

where  $a_h$  is the value obtained from the simple Hertz theory and given by (5).

Resolving the imaginary exponential in (144) into real and imaginary parts, there results

$$\left( \frac{a+x}{a-x} \right)^{i\beta} = \cos \left( \beta \ell n \left| \frac{a+x}{a-x} \right| \right) + i \sin \left( \beta \ell n \left| \frac{a+x}{a-x} \right| \right) \quad (147)$$

It will be noted that  $p(x)$  which should be positive has an infinite number of sign changes as  $x$  approaches  $a$ ,  $-a$ , though it may be that the Hertzian factor  $\sqrt{(a^2 - x^2)}$  has decreased sufficiently where these reversals occur to make the result not more than a mathematical curiosity. Still a solution for which  $p(x)$  stays positive would be of some interest.

No analytical solutions are available for the general rolling contact, tractive or otherwise, of elastically dissimilar cylinders. In reference 6 Bental and Johnson obtained approximate numerical solutions of the integral equations in  $p(x)$ ,  $q(x)$  obtained by substituting in eqs. (115), (118) for  $F$ , ...,  $G'$  from eq. (32). They divided the contact zone into  $N$  ( $=20$ ) equal parts and assumed linear variation of  $p$  and of  $q$  over each interval, whereupon the integrations over each interval could be carried out and the integrals expressed in terms of the values of  $p$  and  $q$  at the points of division. The zones  $L$ ,  $S$  were guessed at; over  $S$  eq. (120) is used and eq. (118) disregarded, thus decreasing both the number of unknowns and equations by the same amount. The resulting linear equations, obtained by applying (115) at all the points of division  $x_i$ , and (118) at the points of division in  $L$ , were then solved for  $p_i$ ,  $q_i$ , and the inequality (119) checked in  $L$  as well as the signs of (120) and (121) in  $S$ . The boundaries of the regions  $L$  were then changed if necessary, and the calculations were repeated until everything checked.

Figure 17, taken from reference 6, shows that for  $\kappa = 0.288$ , corresponding to rolling of aluminum and steel wheels, for non-tractive rolling

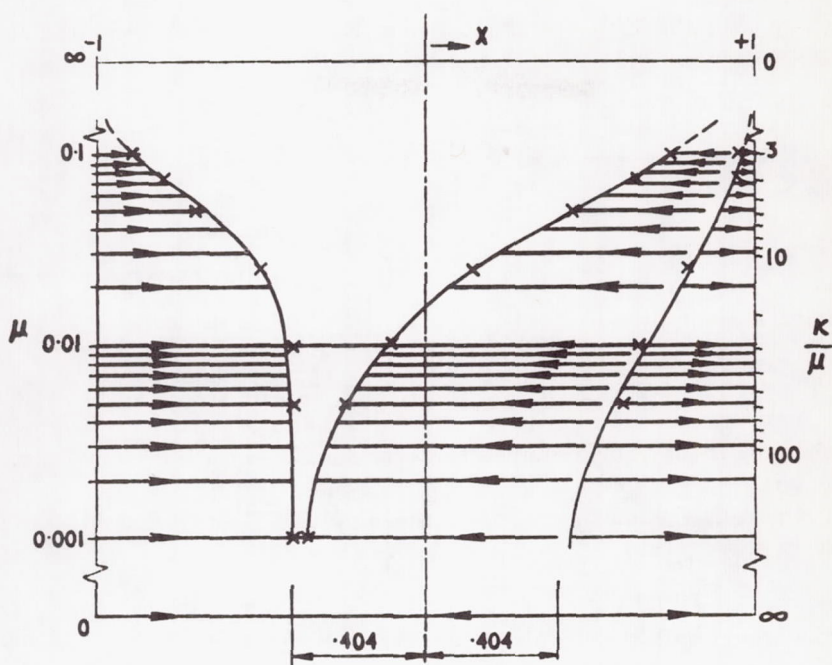


FIGURE 17.—Disposition of slip regions with  $\mu(\kappa = 0.288)$  and  $\kappa/\mu$  (ref. 6).

( $Q=0$ ), there are three slip regions, a leading one,  $S_1$ , a central one,  $S_2$ , and a trailing one,  $S_3$ , with the directions of slip being opposite in successive  $S_i$ . These are separated by two locked regions  $L_1$ ,  $L_2$  with  $q(x)$  changing sign in each one. The region  $L_2$  was too narrow for the size of the interval to be delineated accurately. The creep ratio (necessarily the same) over  $L_i$  was positive showing that the more resilient roller passed the contact faster than its mate. The frictional resistance to rolling has a maximum at  $\mu=0.25$ .

Figure 18, also taken from reference 6, for tractive rolling with  $Q/Q_m=0.75$ , shows two slip regions and a single locked region. In both figures the motion of the wheels through the contact area is to the right. The left half of figure 18 corresponds to the more resilient wheel driving, the right half to its braking. For comparison, the corresponding curves for elastically similar cylinders are also shown in figure 18.

The results of Bental and Johnson are very impressive and show the vast differences that may occur between elastically similar and dissimilar rollers. The existence of slip regions in figure 18, in which the slip is in the opposite direction from that expected from the applied traction, suggests that this situation might arise even for elastically similar rollers—

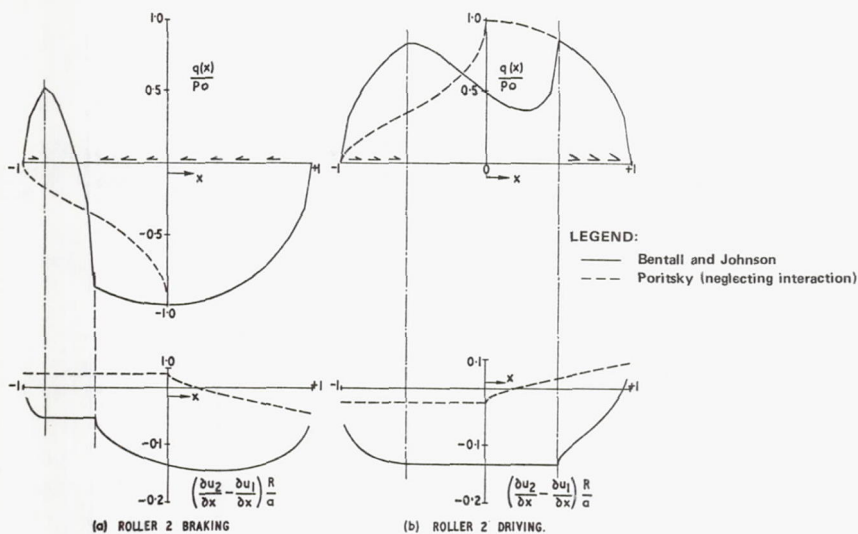


FIGURE 18.—Tractive rolling with slip,  $\kappa = 0.288$ ,  $\mu = 0.05$ ,  $T/\mu N = \pm 0.75$  (ref. 6).

this would lead to a slight decrease in  $Q_m$  for the same creep ratio. These calculations should be repeated with more intervals so as to clear up the nature of the trailing locked region of figure 17, possibly with parabolic rather than linear variations of  $p$  and  $q$  over each interval. These still lead to integrable integrals.

Rubber has frequently been used as one of the rolling elements to demonstrate the existence and shape of the slip regions, and opinion seems to be divided on how representative the results obtained by means of it are. For rubber and metal, Johnson lists the value zero for the fraction  $(1 - 2\nu)/G$  corresponding to rubber (see table II, ref. 4). This is based on using the Poisson ratio  $1/2$  for rubber. Such a value of  $\nu$ , used in eq. (53), would dissociate the interaction of  $q(x)$  on the normal displacement  $n_2$  and of the normal pressure  $p(x)$  on  $t_2$  and would reduce  $\alpha$  in eq. (116) to  $(1 - 2\nu_1)/2(1 - \nu_1)$  (though not to zero thus eliminating all interactions between  $p$  and  $q$ ). However, while for rubber  $\nu$  is close to  $1/2$ , it is also true that for rubber the Young's modulus is very small in comparison with that of metals. Indeed, the volume modulus for rubber, which is proportional to  $E/(1 - 2\nu)$ , may be, for some rubbers, of the same order of magnitude as for metals.

##### 5. SLIP AND CREEP OVER NONCYLINDRICAL CONTACTS

By noncylindrical contacts is meant contacts of noncylindrical bodies or of cylinders with nonparallel axes, where the contact area is an ellipse (or circle). Relatively few analytical solutions similar to those in the



second through fourth sections are available, and some of these will be described in the sixth section. In this section we present a very incomplete review of some of the approximate theoretical calculations of slip and creep over the elliptic contact areas and of experimental results that have appeared over the last 20 years.

For rolling contacts we assume the direction of rolling of the upper body over the lower one to be that of positive  $x$  with velocity  $U$  (see figs. 1 and 2); relative to a system of axes in which the contact zone is fixed the two bodies move to the left with velocity  $-U$ .

More parameters are now involved in the contact phenomena than in the case of cylindrical contacts: the two elliptic semi-axes; the inclination of the major semi-axis toward the  $x$ -axis; the two components of tangential force  $F_x, F_y$  transmitted across the contact, as well as the moment  $M_z$  of the tangential tractions about the common normal at the center of the ellipse (the  $z$ -axis); or alternatively, the creep ratios  $\xi_x, \xi_y$  of the upper body (at infinite  $z$ ) relative to the lower one in the  $x$ - and  $y$ -directions, and  $\omega_z$ , the rate of spin about the common normal. Moreover, the shape of the dividing curve  $\Gamma$  between  $L$ , the locked, and  $S$ , the slip area has to be determined.

Different conditions apply over  $S$  and over  $L$ . Over  $S$  slipping takes place, the resultant of the shearing stress  $(X, Y)$  is equal to its limiting value  $\mu p$ , and it is directed in the direction of the slip of body 2 over body 1. Over  $L$  the resultant shear is less than its limiting value, and no slip occurs. The calculation of the slip velocity is obtained by adding the velocity of relative motion of body 1 over body 2 as rigid bodies, and the velocities due to the elastic displacements. These items will be discussed in more detail in the sixth section.

An approximate solution for the rolling of a sphere of diameter  $D$  on a plane under a normal force  $P$  and tangential contact force  $F_x$  was obtained by Johnson (ref. 12). The Hertz theory shows that contact occurs over a circle of radius  $a$  given by

$$a^3 = \frac{3(1-\nu)PD}{8G} \quad (148)$$

and that the pressure distribution is

$$p = \frac{3P}{2\pi a^3} (a^2 - r^2)^{1/2}, \quad r < a \quad (149)$$

If the tractive force  $F_x$  is equal to its maximum possible value  $\mu P$ , then slipping takes place over the entire circle of contact and the distribution of traction over it is given by

$$X = \mu p = \frac{3\mu P}{2\pi a^3} (a^2 - r^2)^{1/2}, \quad r < a \quad (150)$$

It has been shown by Cattaneo (ref. 13) that if from (150) is subtracted the traction

$$X' = \frac{3\mu P}{2\pi a^3} [(a')^2 - r^2]^{1/2}, \quad r < a', \quad a' < a \quad (151)$$

whose resultant is  $\mu P(a'/a)^3$ , then the condition of constant displacement

$$u = \text{const.}, \quad v = 0 \quad (152)$$

is approximately satisfied over the interior of  $C'$ , the circle  $r < a'$ , while outside  $C'$  and inside the contact circle  $C$ :  $r < a$  the slip condition  $X = \mu p$  holds. By adding a constant translation to the upper solid which is equal and opposite to (152) there results a solution with zero displacements over  $C'$  and the slip condition between  $C'$  and  $C$ .

For the lower solid the net  $u$ -displacements are obtained by reflection in the plane  $z=0$  and changing sign. The solution obtained represents a *stationary contact* between a sphere and a plane, which, in addition to the pressure force  $P$ , is also transmitting the tangential force

$$F_x = \left[ 1 - \left( \frac{a'}{a} \right)^3 \right] \mu P \quad (153)$$

with a locked area  $L$  over the circle  $C'$  over which no displacement has taken place, and a ring-shaped slip area  $S$ :  $a' < r < a$  over which slip has occurred.

Johnson obtained an approximate solution for *rolling contact* between the sphere and the plane by shifting the circle  $C'$  and the stress distribution (151) to the right so that  $C'$  touches the circle  $C$  at its leading point  $x=a$ ,  $y=0$ , so that its center is moved to

$$x = c = a - a', \quad y = 0 \quad (154)$$

From eq. (153) follows

$$\frac{c}{a} = \frac{a - a'}{a} = 1 - \frac{a'}{a} = 1 - \left( 1 - \frac{F_x}{\mu P} \right)^{1/3} \quad (155)$$

With the shift in  $C'$  there results a constant strain difference in  $e_x = \partial u / \partial x$  between the upper and lower bodies. Hence the sphere now creeps over the plane with the creep ratios

$$\xi_x = \frac{3\mu P(4-3\nu)c}{16Ga^2}, \quad \xi_o = 0 \quad (156)$$

and substituting from eqs. (148), (155)

$$\xi_x = \frac{\mu(4-3\nu)}{2(1-\nu)} \frac{a}{D} \left[ 1 - \left( 1 - \frac{F_x}{\mu P} \right)^{1/3} \right] \quad (157)$$

It turns out, however, that over the area outside  $C'$  and to the right of  $x=c$  the slip is in the wrong direction.

Figures 19 and 20 show, for  $a'/a = \frac{1}{2}$ , the locked area  $L$  for a stationary and rolling contact, respectively. The shaded area in figure 20 is where the slip is in the wrong direction.

Johnson (ref. 12) also obtains a solution for a rolling sphere with a tangential force  $F_y$  normal to the direction of rolling. With  $a'$  defined by

$$F_y = \left[ 1 - \left( \frac{a'}{a} \right)^3 \right] \mu P \quad (158)$$

and  $Y$  given by the difference of the right-hand members of (150) and (151), the stationary contact case results; with  $C'$  moved as above the rolling contact case is obtained. The creep is obtained as

$$\xi_x = 0, \quad \xi_o = \frac{\mu(4-\nu)}{2(1-\nu)} \frac{a}{D} \left[ 1 - \left( 1 - \frac{F_y}{\mu N} \right)^{1/3} \right] \quad (159)$$

To verify the above relations, experiments were carried out on hard steel balls rolling on a hard steel slab. The equipment used is shown in figure 21, taken from reference 12. Two balls were fastened to a spindle and were loaded by means of two hangers mounted on ball bearings at each end of the spindle. With the width dimension of the slab maintained horizontal and the length tilted at an angle  $\alpha$  to the horizontal (see fig. 21a), the ratio of the tangential to normal force  $F_x/P$  becomes equal to  $\tan \alpha$  and eq. (157) yields

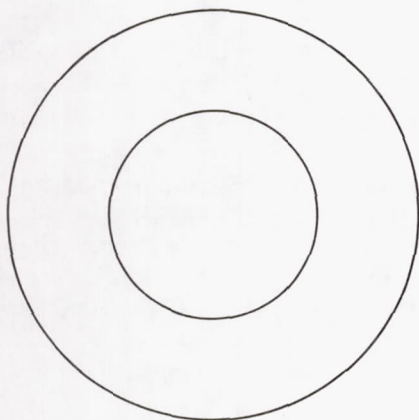


FIGURE 19.—Illustrating Cattaneo's slip and locked regions for contact of sphere and plane.

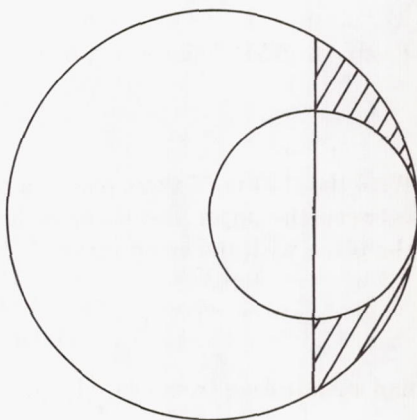


FIGURE 20.—Illustrating Johnson's solution for slip and locked regions for rolling contact of sphere and plane.



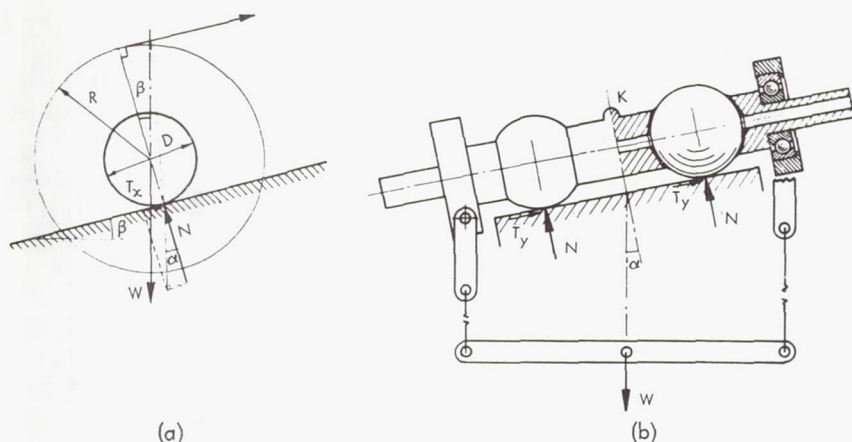


FIGURE 21.—Experimental apparatus for measuring creep under action of tangential forces. (a) Application of a longitudinal tangential force. (b) Application of a transverse tangential force (ref. 12).

$$\xi_x = \frac{\mu(4-3\nu)}{2(1-\nu)} \frac{a}{D} \left[ 1 - \left( 1 - \frac{\tan \alpha}{\mu} \right)^{1/3} \right] \quad (160)$$

The proportionality of  $\xi_x$  to  $a$  (varied by changing the weight) was strictly verified for each value of  $\alpha$ . A plot of  $\xi_x D/a$  vs  $\tan \alpha$  showed good agreement with eq. (160) for small values of  $\tan \alpha$ , but discrepancies occurred as  $\tan \alpha$  approached the value  $\mu$ , estimated from the point at which the slope of the curve becomes infinite as  $\mu = 0.09$ . These discrepancies are ascribed by Johnson to the wrong direction of slip over the shaded area of figure 20.

By maintaining the long direction of the slab horizontal, inclining its width (see fig. 21b) at an angle  $\alpha$ , and allowing the balls to roll a distance  $L$  and return, the effects of a normal force  $F_y = P \tan \alpha$  on the deflection in the normal direction were obtained and  $\xi_y$  evaluated and compared with eq. (159), written in a form similar to (160) but with  $(4-3\nu)$  replaced by  $(4-\nu)$ . Again, very good agreement was obtained for small values of  $F_y/P$  with  $\mu$  estimated from the point where the slope of  $\xi_y D/a$  vs  $\tan \alpha$  became infinite, as  $\mu = 0.11$  for dry contact,  $\mu = 0.09$  for lubricated contacts; while discrepancies resulted for  $0.04 < \tan \alpha < \mu$ , with the measured  $\xi_y$  exceeding the theoretical values.

In a companion paper (ref. 14), Johnson studied the effects of spin  $\omega_z$  about the normal axis (the  $z$ -axis) of a rolling ball under a normal load  $P$  (but without tangential forces). He had discovered accidentally that in a simple thrust bearing made up of two concentrically rotating flat disks, rotating at different speeds, with evenly spaced balls between

them whose centers were at the same distance  $R$  from the axis, the balls did not move around in a circular path but exhibited a creeping motion outward, and this occurred even when the centrifugal forces were negligible. By postulating a proper superposition of tractive forces  $X$ ,  $Y$ , Johnson was able to compute a ball creep at right angles to the direction of rolling, generated by the spin  $\omega_z$  of the balls caused by their contacts with the disks. Assuming that there is no slip over the entire circle of contact he arrives at the transverse creep given by

$$\xi_y = \frac{2(2-\nu)}{3(3-2\nu)} \frac{\omega_z a}{U} \quad (161)$$

and obtains the spin moment

$$M_z = -Ga^4\omega_z/U \quad (162)$$

These results are presumably valid only when  $\xi_y$ ,  $\omega_z$  are very small. Johnson suggests that when slip does occur over part of the contact circle, eqs. (161), (162) should be replaced respectively by

$$\frac{\xi_y D}{\mu a} = f_1 \left( \frac{D}{\mu R} \right) \quad (163)$$

$$\frac{M_z}{\mu Pa} = (1-\nu)f_2 \left( \frac{D}{\mu R} \right) \quad (164)$$

To check the above relations observations were made on three balls between two disks, the lower one stationary, the upper one loaded with a weight and capable of being rotated, with the ball centers at a distance  $R$  from the axis of rotation and  $120^\circ$  apart. The upper disk was rotated slowly until one complete circulation of each ball was obtained, and  $\delta$ , the radial movement of the ball was measured with a micrometer, while  $\xi_y$ ,  $\omega_z a/U$  were obtained from

$$\xi_y = \delta/2\pi R, \quad \omega_z a/U = a/R \quad (165)$$

While eq. (161) was checked for various values of  $a$  and ball diameters  $D$ , it was difficult to correlate the results with the form (163) in view of the uncertainty in the coefficient of friction  $\mu$  for slip. Measurements of the coefficient of sliding friction did not show much variation, but the measured-values of  $\xi_y$  required assumption of wide variations of  $\mu$  depending upon surface finish and hardness of the track material—the balls were made of hard steel and were polished—for correlation of results into the form (163). For this reason  $\mu$  was taken out of (163) and  $\xi_y D/a$  was plotted vs  $D/R$ . For  $D/R < 0.1$ , all the points fell on a straight line agreement with (161); but for larger values of  $D/R$ ,  $\xi_y D/a$  levelled off to constant values in a manner depending upon the surface finish and hardness of the race and whether the contacts were lubricated or not.

The effects of simultaneous tangential forces and spin were studied by Johnson (ref. 15) both analytically and experimentally. For small  $\omega_z$  and  $F_y$ , eqs. (159), (161) yield, by superposition, for  $\nu=0.3$ ,

$$\frac{\xi_y D}{a} = 0.47 \frac{\omega_z D}{U} - 0.88 \frac{F_y}{P} \quad (166)$$

Careful measurements led to a family of curves satisfying the empirical equation

$$\frac{\xi_y D}{a} = 0.44 \frac{\omega_z D}{U} - 3.3 \left( \frac{\omega_z D}{U} \right) \left( \frac{F_y}{P} \right) - 1.10 \left( \frac{F_y}{P} \right) \quad (167)$$

provided that  $(\omega_z D/U) < 0.12$ ,  $(F_y/P) < 0.35$ . The product term indicates a slight interaction between the effects of  $F_y$  and  $\omega_z$  in producing creep.

Measurements were also carried out for large  $F_y$  and  $\omega_z$ , where slip must occur over part of the circle of contact. The results are shown in figure 22, where  $r$  is used for the radius of the ball and  $T_y$  for  $F_y$ .

As a result of the transverse creep induced by the spin, a ball rolling between two races may depart from the position in which it makes contact with the races at opposite ends of its diameter, and tend to wedge itself in

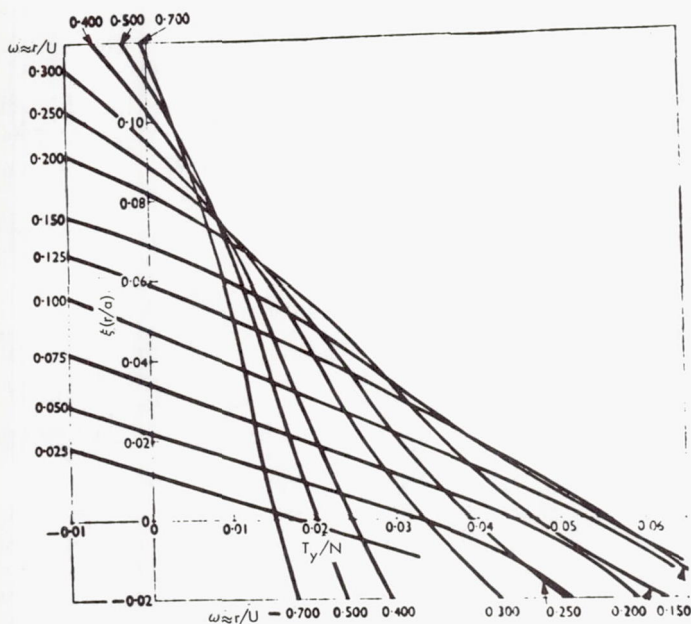


FIGURE 22.—Creep measurements with larger tangential forces and spin from the thrust bearing apparatus (ref. 15).



so that its points of contact are not diametrically opposite, and the forces exerted by the races on the ball have large tangential components which oppose further ball creep. The resulting positions of equilibrium can be predicted roughly from eqs. (166), (167) or the curves of figure 22 by equating the creep to zero.

Johnson also calculated the moment  $M_z$  about the spin axis, and carried out frictional resistance measurements to rolling with spin.

Turning to general, elliptic contact area of rolling bodies, Vermeulen and Johnson (ref. 16) studied contacts transmitting both normal and tangential forces, assuming the locked region  $L$  to be an ellipse similar to the contact ellipse and tangent to it at the foremost point.

A method first suggested by Haines is the strip method; it is useful when the ellipse of contact is narrow in the rolling direction

$$x^2/a_o^2 + y^2/b_o^2 = 1, \quad b_o \gg a_o \quad (168)$$

This method consists in breaking up the ellipse (168) into strips parallel to the  $x$ -axis of width  $dy$ , and applying to each strip the results of the third section, with a locked area  $L$  and a slip area  $S$ , a normal load obtained from the Hertzian ellipsoidal pressure distribution, and a creep ratio determined from the relative motion of the contacting solids.

As an illustration we follow Johnson (ref. 4) in considering the Heathcote slip, i.e., the motion of a ball of radius  $R$ , rolling freely in a straight conforming groove whose cross section by planes  $x = \text{const}$  is a circular arc of radius  $\rho$  close to  $R$ . The ball is loaded normally by means of a plate with a force  $P$ , and the plate is moved in direction of positive  $x$  with velocity  $2U$ . The Hertzian ellipse of contact (168) of semi-axes  $a_o$ ,  $b_o$ , will be narrow in the rolling,  $x$ -direction, and due to the conformity of the groove, long in the transverse,  $y$ -direction, where it is wrapped on the surface of the sphere,\* so that the effects due to the variable height  $z$  of the ellipse

$$z = \frac{x^2 + y^2}{2R} \sim \frac{y^2}{2R} \quad (169)$$

must be taken into account. Heathcote pointed out that, since the ball can only pivot at any instant about a fixed axis, say the line cutting the sphere in

$$x = 0, \quad y = \pm \gamma b_o, \quad z = z_o = \frac{\gamma^2 b_o^2}{2R}, \quad 0 < \gamma < 1 \quad (170)$$

the strips in  $\gamma b_o < |y| < b_o$  must slip on the groove in one direction, the strips  $|y| < \gamma b_o$  in the opposite direction as shown in figure 23, with the

---

\*More precisely, on a surface of curvature  $1/2(1/R + 1/\rho)$ .

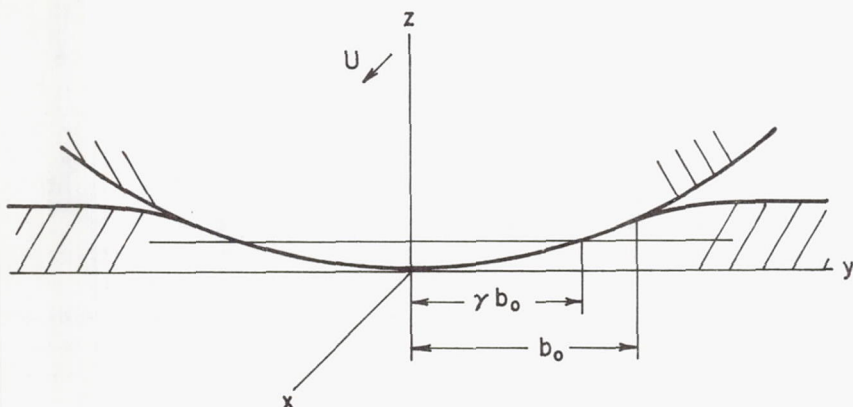


FIGURE 23.—Rolling of sphere in a groove.

friction force  $f_x = \pm \mu p(x, y)$  reversing sign at  $y = \pm \gamma b_0$ . This leads to  $\gamma = \frac{1}{2}$  and yields the rolling resistance

$$F = 0.080 \mu P b_0^2 / R^2 \quad (171)$$

The above calculation takes no account of the tangential displacements produced by the friction forces. Johnson proceeded to take the tangential strains into account by postulating for each strip of width  $dy$  a (possible) locked region  $L$  located on the forward end of the strip (see fig. 4), a slip region  $S$  back of it with a shear force distribution over the strip as indicated in figure 4, and a creep ratio determined by the creep of the ball as a whole over the groove and the added creep for each strip due to its height above the axis of rotation.

Figure 24 shows, for  $R/\rho = 0.948$  and  $0.815$ , the shape of the ellipse of contact, the locked and the slip areas, and the direction of slip (for a direction of rolling velocity opposite to the one used in this paper). The ratios of the friction force to the Heathcote value (171) are  $0.824$  and  $0.558$ , respectively.

Halling (ref. 17) applied a similar technique to the study of the effects of conformity on the stick and slip phenomena between rolling elements revolving about parallel horizontal lines; the upper, spherical body, rolling along a hollowed-out, conforming, track provided by the lower one, with a driving moment applied to the upper one. Figure 25 gives an example of the shape of the locked area. Two locked areas, symmetrically situated about  $y = 0$ , start at points corresponding to the  $y$ -values  $\pm \gamma b_0$  of the rotation axis, on the forward side of the contact ellipse, and widen but do not fuse toward the lee side. For other values of the parameters involved, the two locked areas may fuse into a single area.

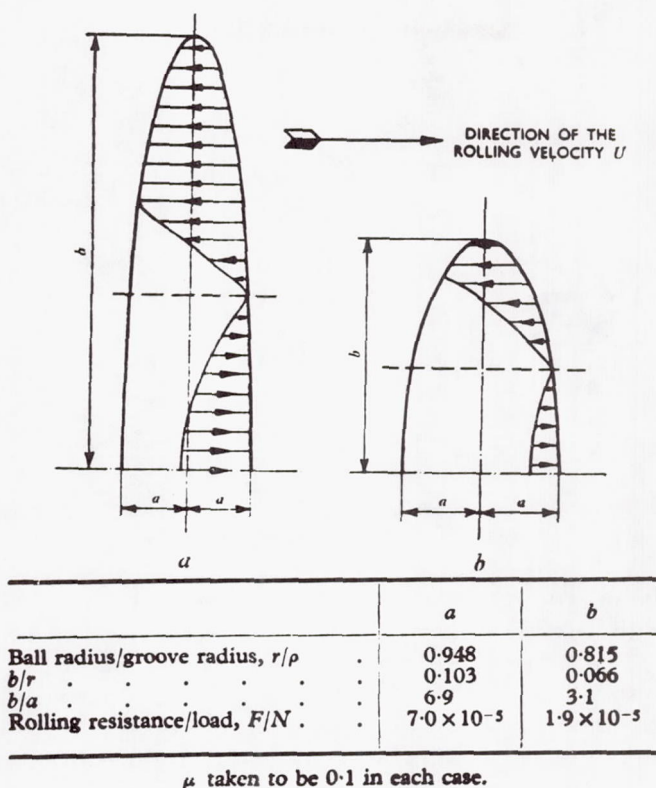


FIGURE 24.—Calculated slip on the contact area of a ball rolling freely in a conforming straight groove (ref. 15).

Halling also examined wheel-rail systems with a hollowed-out track on the rail and a convex roller, or a convex rail shape and a concave roller shape.

In references 18 and 19 results on creep measurements and on wear due to fretting over the slip areas are given which confirm the theory. The experiments were carried out by using as a rolling element a ball which had been machined out, rendered radioactive in a nuclear pile, and mounted on a copper mandrel fixed on a horizontal spindle held in ball bearings. The ball specimen rested on the track of a wheel fixed on the driven shaft, also horizontal, and provision was made for applying a measured load between the rolling element and the wheel. The creep rate was measured by reflecting beams of light from mirrors attached to the rolling element spindle and the driving wheel axle onto a graduated curved scale.

During the tests the ball specimen was driven by the track against the



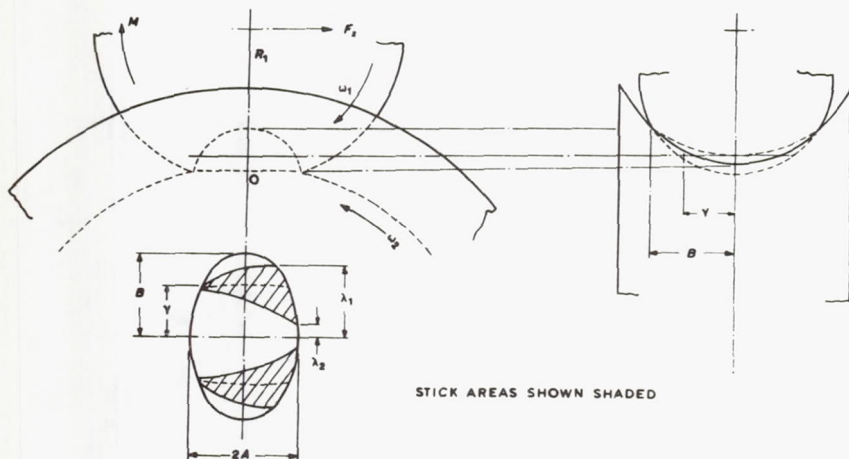


FIGURE 25.—Details of the contact geometry where a degree of conformity exists between the two rolling elements (ref. 17).

resistance offered by the supporting ball bearings; this resisting force had previously been determined for various normal forces applied to the uncoupled ball. Curves of the creep rate vs the normal load were plotted for rolling in grooves of different curvatures, and these compared favorably with the theoretically predicted curves.

The wear rates were obtained in two ways. The rolling element was made of softer steel than the hardened steel wheel track, so that the wear occurred mostly on the ball. The amount of wear was obtained by radioactive measurements of the track, as well as by weighing the ball. The weight of the material transferred increased with the rolling distance up to a certain length of rolling, then levelled off indicating that some of the material that had been transferred to the track was beginning to be transferred back to the wheel.

Empirical equations are given in references 18 and 19 for the rate of wear of the power law form

$$\dot{w} = kp^n v^m \quad (172)$$

where  $p$  is the pressure,  $v$  is the local rate of slip,  $k$ ,  $n$ ,  $m$  are constants;  $k$  depends on the surface finish of the track as well as its hardness (which was always greater than the hardness of the ball).

Halling (ref. 20) studied the microslip between a rolling ball and its conforming circular track in a thrust ball bearing. The ball is now subjected to forces in the plane of motion, as well as to an applied spin moment about the normal to the contact area. There results now a considerable lack of symmetry of the locked areas about the minor axis of the ellipse of contact when the load is axial as shown in figure 26. For a



radially loaded bearing the locked regions are symmetric, and the regions of wear are indicated in figure 27.

It is known that microslip is not the only factor that contributes to resistance to rolling. Another source of rolling resistance is that due to internal damping of the material of the ball and race, which shows up as hysteresis between the stress and strain at loading and unloading even in the elastic range at low stresses, and in more substantial hysteresis losses in the volume under the loaded area for the high stresses that occur in ball bearings. These losses have become familiar since the work of Drutowski (see, for instance, ref. 4, pp. 113-131) who used a ball rolling between a stationary lower plate and a moving upper plate, and measured the force that had to be applied to the moving plate for different normal loads, ball diameters, etc. In a more recent paper Drutowski (ref. 21) carried out similar tests for a ball rolling in a conforming groove in the lower plate with an elliptic area of contact, and compared the rolling resistance with that of a crowned roller rolling on a flat plate where the Hertz ellipse had the same semi-axes as the ellipse for the ball-groove combination. Presumably the hysteresis resistance to rolling was the same in the two cases, and the difference represented the added resistance due to microslip. Figure 28 shows a typical case.

In some cases, such as a ball in an angular contact bearing, it is kinematically impossible for the ball to roll relative to both the internal and the external races; in such cases the ball will roll on the race for which the ellipse has a higher eccentricity, and will roll and spin relative to the other race. Based on rigid body spin rotations, the friction torques for all the balls of a bearing were calculated by Poritsky et al. (ref. 22) both for a thrust (axial) load and for a load normal to the axis. The effect of the compliance of the ball and races remains to be included as well as possible transverse creep due to spin that may cause the ball to get



FIGURE 27.—Wear track for a radially loaded ball bearing together with predicted microslip pattern for this condition (ref. 20).



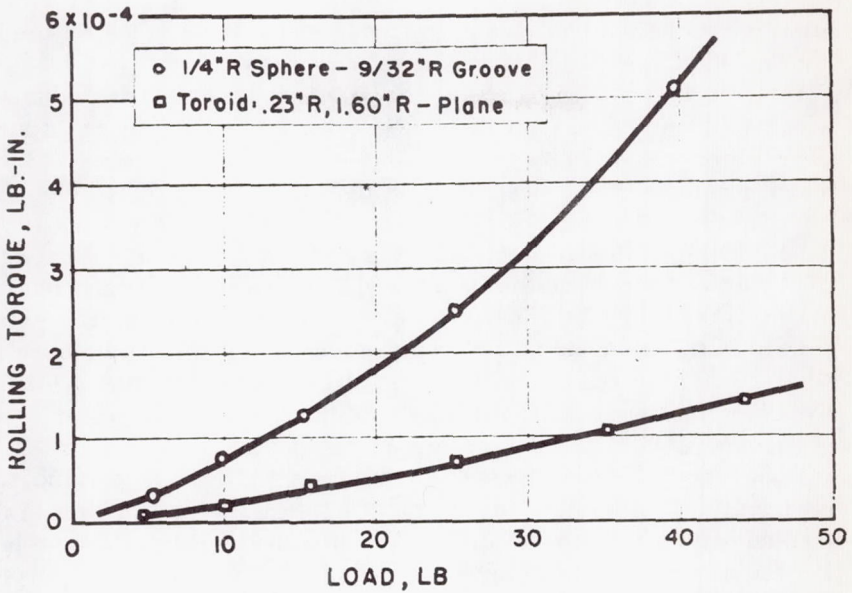


FIGURE 28.—Comparison of total torque for equivalent contact ellipses having a ratio of major to minor diameters of approximately 3.8 (ref. 21).

wedged in so that its contacts with the races are not diametrically opposite. For high-speed bearings the effect of centrifugal forces causes the contacts to move even further from diametrically opposite positions, whereupon the angular velocity along the line through the centers of the two contact areas, resolved at each contact, has a rolling as well as a spinning component. Gyroscopic moments of the ball contribute tangential forces normal to the rolling velocity.

Measurements of friction forces for balls rolling in a groove were carried out by Reichenbach (ref. 23). He repeated and checked the Drutowski measurements for a ball rolling between two horizontal plates, then replaced the lower plate by straight-sided V-grooves making various angles with each other, and later by grooves with curved sides and several degrees of conformity with the ball radius. The results essentially agreed with the calculations in reference 22. The resistance to rolling due to spin moment turned out to be many times larger than the resistance due to hysteresis. While the coefficient of friction (i.e., the ratio of the tangential force on the moving plate, per ball, to the normal force) for a ball-bearing is of the order 0.001 to 0.003, Reichenbach points out that it is of great concern to designers. In a jet engine it may result in heavy power loss whose dissipation becomes an important item; it is of even greater concern in supersonic jet engine planes. While the friction forces

are small in inertial guidance gyroscopic systems, unless their nature is known and allowed for, they may cause cumulative errors.

Dietrich, Parker, and Zaretsky (ref. 24) carried out tests for spinning a ball (without rolling) revolving about a fixed axis relative to a stationary groove, conforming to ball, and resulting in elliptic area of contact under a normal load. Under unlubricated contact conditions, they obtained friction torques that led to coefficients of friction varying from 2.6 to 6.0. For lubricated contacts, the coefficient of friction decreased from 0.5 to 0.55 with increasing load and Hertzian stress. It was pointed out in a discussion of this paper by Flom et al. (ref. 25) that the stress pattern in the revolving ball, though stationary in space, revolves backwards relative to the material of the ball, so that losses due to elastic hysteresis must occur in it. Yet it is doubtful whether the high coefficient of friction can be accounted for by hysteresis effects alone.

Haines and Ollerton (ref. 26) and Haines (ref. 27) studied photoelastically the stresses over circular contact areas, and applied the strip method to obtain the shape of the locked area and the stress distribution. For the case of slip in the direction of rolling only, they obtained a lemon-shaped area for  $L$ .

Kalker (ref. 28) proposes an empirical formula for the tangential forces transmitted across an elliptic contact area in rolling contact with creepage. In reference 29 he extends the strip theory to general creepage and spin.

#### 6. ANALYTICAL SOLUTIONS OF THREE-DIMENSIONAL CONTACT PROBLEMS

In this section we review the Hertz solution and the Mindlin solution for a stationary contact from point of view of the analogies to classical problems in potential theory. We also present some more recent analytical solutions by de Pater, Kalker, and Mow et al. No complete analytical solutions for rolling contacts with both slip and locked area are available. We also present the Hertz and Cattaneo solutions for the displacements produced in a semi-infinite solid by the application of a normal and tangential force at a point on the boundary. These form the kernels of the integral equation methods which will be discussed in the next section. A brief introduction to ellipsoidal harmonics is also included.

A vertical, concentrated normal pressure force  $P$ , applied at the origin  $O$  to the boundary  $z=0$  of the semi-infinite solid  $z>0$ , produces in it the displacements (ref. 30)

$$u = \frac{P}{4\pi G_1} \left( \frac{xz}{R^3} - (1-2\nu_1) \frac{x}{R(R+z)} \right)$$

$$v = \frac{P}{4\pi G_1} \left( \frac{yz}{R^3} - (1-2\nu_1) \frac{y}{R(R+z)} \right)$$

$$w = \frac{P}{4\pi G_1} \left( \frac{z^2}{R^3} + \frac{2(1-\nu_1)}{R} \right) \quad (173)$$

where  $R = (x^2 + y^2 + z^2)^{1/2}$ . These displacements may be expressed in the form

$$\mathbf{v} = (u, v, w) = \frac{1}{4\pi G_1} \left[ -\nabla \left( \frac{\partial \psi}{\partial z} \right) - (1-2\nu_1) \nabla \chi + 2\mathbf{k}(1-\nu_1) \nabla^2 \psi \right] \quad (174)$$

where  $\mathbf{k}$  is a unit vector in direction of increasing  $z$ , and

$$\psi = PR, \quad \chi = P \ln(R+z), \quad \nabla^2 \psi = 2P/R \quad (175)$$

It will be noted that  $\psi$  is biharmonic; it can be interpreted as the potential of the field of a central force of constant magnitude, directed to 0 and independent of  $R$ ;  $\chi$  is harmonic except on the negative  $z$ -axis; it represents the potential of a uniform charge distribution over the negative  $z$ -axis.

The vector displacement may be put in the form

$$\mathbf{v} = \frac{P}{4\pi G_1 R} \{ \mathbf{u}_r \sin \theta [\cos \theta - (1-2\nu_1)/(1+\cos \theta)] + \mathbf{k}(\cos^2 \theta + 2(1-\nu_1)) \} \quad (176)$$

where  $\mathbf{u}_r$  is the unit vector in direction of increasing  $r = (x^2 + y^2)^{1/2}$  and  $\cos \theta = z/R$ ; this shows that  $\mathbf{v}$  is axially symmetric about the  $z$ -axis.

The displacement components in (173) are homogeneous in  $x, y, z$  of degree  $-1$ . Hence the strain and stress components are homogeneous of degree  $-2$ . In particular, the tractions across a hemisphere of constant  $R$  are given by

$$\mathbf{F}_R = \frac{P}{2\pi R^2} \left[ \mathbf{u}_R \left( 3 \cos \theta - \frac{1-2\nu_1}{1+\cos \theta} \right) + \mathbf{k} \frac{(1-2\nu_1) \cos \theta}{1+\cos \theta} \right] \quad (177)$$

where  $\mathbf{u}_R$  is a unit vector in direction of increasing  $R$ . Alternative forms are obtained by resolving  $\mathbf{u}_R$  first along  $\mathbf{u}_r, \mathbf{k}$ , then expressing  $\mathbf{k}$  (in 177) in terms of  $\mathbf{u}_R, \mathbf{u}_\theta$ , the unit vector in direction of increasing  $\theta$

$$\mathbf{F}_R = \frac{P}{2\pi R^2} \left[ \mathbf{u}_r \cos \theta \left( 3 \cos \theta - \frac{1-2\nu_1}{1+\cos \theta} \right) + \mathbf{k} 3 \sin \theta \cos \theta \right] \quad (177')$$

$$\mathbf{F}_R = \frac{P}{2\pi R^2} \left[ \mathbf{u}_R \left( 3 \cos \theta - \frac{(1-2\nu_1) \sin^2 \theta}{1+\cos \theta} \right) - \mathbf{u}_\theta \frac{(1-2\nu_1) \cos \theta \sin \theta}{1+\cos \theta} \right] \quad (177'')$$

The form (177') is useful in obtaining the resultant traction across the hemisphere

$$\frac{\mathbf{k}P}{2\pi R^2} \int_0^{\pi/2} 2\pi R^2 \cdot 3 \cos^2 \theta \sin \theta d\theta = \mathbf{k}P \quad (178)$$



From the form (177'') it will be noted that the traction across a sphere of constant  $R$  consists of a normal stress component  $p_{RR}$  and a shear component  $p_{R\theta}$  equal respectively to the negatives of the net coefficients of  $u_R, u_\theta$  in (177''); moreover,  $p_{RR}$  is compressive inside a certain cone, tensile outside. There is thus quite a difference between the way the applied load  $P$  is carried over the aspherical surface in this case and the manner in which it is transmitted in the two-dimensional case over the cylindrical surfaces of constant  $r$ , where it is carried purely as a normal stress  $p_{rr}$ . The nearest approach to a radial traction occurs across horizontal planes of constant  $z$ , across which the traction vector is

$$F_z = \frac{3P \cos^2 \theta}{2\pi R^2} u_R \quad (179)$$

which really represents both a normal component as well as a shear component.

Over the boundary  $z=0$  it follows from (179) that the traction vanishes (except at the origin), while (177') yields

$$v = \frac{P}{4\pi G_1 R} [-u_r(\frac{1}{2} - \nu_1) + k(3 - 2\nu_1)] \quad (180)$$

Hence follow the strain components

$$e_r = \frac{\partial v}{\partial r} = \frac{P(1 - 2\nu_1)}{8\pi G_1} = \frac{1}{R^2}, \quad e_\phi = \frac{v_r}{R} = -e_r \quad (181)$$

respectively, tensile and compressive. Since  $p_{zz}$  vanishes, it follows that

$$e_z = 0 \quad (182)$$

and that

$$p_{rr} = -p_{\phi\phi} = \frac{P(1 - 2\nu_1)}{4\pi G_1} \frac{1}{R^2} \quad (183)$$

Applying superposition to a distributed pressure over an area  $A$ , we obtain for the displacements (174), where now  $\psi, \chi$  are given by

$$\psi(P) = \iint_A p(P') R(P, P') dP' \quad (184)$$

$$\chi(P) = \iint_A p(P') \ell n[z + R(P, P')] dP'$$

and  $P: (x, y, z)$  is an arbitrary field point,  $P': (x', y', 0)$  a point in the area  $A$ ,  $dP'$  the element of area of  $A$  over which area the integration is carried out.

Alternatively the superposition may be carried out directly on (173) by replacing  $x, y, z, R$  by  $x-x', y-y', z, R(P, P')$ , replacing  $P$  by  $p(x', y') dx' dy'$ , and integrating over  $A$ . In particular, for  $z=0$ , there results

$$\begin{aligned} u(P) &= -\frac{(1-2\nu_1)}{4\pi G_1} \iint_A \frac{p'(x', y') (x-x')}{R(P, P')} dP' \\ v(P) &= -\frac{(1-2\nu_1)}{4\pi G_1} \iint_A \frac{p'(x', y') (y-y')}{R^2(P, P')} dP' \\ w(P) &= \frac{2(1-\nu_1)}{4\pi G_1} \iint_A \frac{p'(x', y')}{R(P, P')} dP' \end{aligned} \quad (185)$$

From (185)<sub>3</sub> follows the analogy between  $w(P)$  and  $p(P')$ , and the gravitational potential  $V(P)$  at  $P$  of a mass distribution  $\sigma(P')$  over  $A$ , proportional to  $p(P')$ , expressed by

$$w|_{z=0}: V = \text{const } p:\sigma \quad (186)$$

By reflecting the axes and the solution of eqs. (173) to (185) in the plane  $z=0$  and replacing the elastic constants by  $G_2, \nu_2$ , one obtains a similar solution for the half space  $z<0$ .

The displacements and stresses described by eqs. (173) to (183) are also of interest for distributed pressures, where they describe the asymptotic behavior of the solutions (174), (184), (185) at infinity, provided the resultant pressure does not vanish, and the origin is placed at the centroid of the pressure distribution.

The solution (173) is due to Boussinesq. We next introduce the solution, due to Cerruti, of the elasticity equations for the displacements in  $z>0$  produced by the application of a concentrated shearing force  $F_x$ , applied at the origin  $O$  (ref. 30)

$$\begin{aligned} u &= \frac{F_x}{4\pi G_1} \left[ \frac{1}{R} + \frac{x^2}{R^3} + (1-2\nu_1) \left( \frac{1}{z+R} - \frac{x^2}{R(z+R)^2} \right) \right] \\ v &= \frac{F_x}{4\pi G_1} \left[ \frac{xy}{R^3} - (1-2\nu_1) \frac{xy}{R(R+z)^2} \right] \\ w &= \frac{F_x}{4\pi G_1} \left[ \frac{xz}{R^3} + (1-2\nu_1) \frac{x}{R(R+z)} \right] \end{aligned} \quad (187)$$

The displacements (187) can also be expressed in terms of a single harmonic function  $F_1$

$$F_1(x, y, z) = F_x \Omega = F_x [z \ell n(R+z) - R] \quad (188)$$

as follows:

$$\begin{aligned}
 u &= \frac{1}{2\pi G_1} \left[ \frac{\partial^2 F_1}{\partial z^2} + \nu_1 \frac{\partial^2 F_1}{(\partial x)^2} - \frac{z}{2} \left( \frac{\partial^3 F_1}{(\partial x)^2 \partial z} \right) \right] \\
 v &= \frac{1}{2\pi G_1} \left[ \nu_1 \frac{\partial^2 F_1}{\partial x \partial y} - \frac{z}{2} \frac{\partial^3 F_1}{\partial x \partial y \partial z} \right] \\
 w &= \frac{1}{4\pi G_1} \left[ (1 - 2\nu_1) \frac{\partial^2 F_1}{\partial x \partial z} - z \frac{\partial^3 F_1}{\partial x (\partial z)^2} \right]
 \end{aligned} \tag{189}$$

Differentiation of  $F_1$  in (188) yields

$$\begin{aligned}
 \frac{\partial}{\partial z} [z \ell n(R+z) - R] &= \ell n(R+z) \\
 \frac{\partial^2}{\partial z^2} [z \ell n(R+z) - R] &= \frac{1}{R}
 \end{aligned} \tag{190}$$

As pointed out following eq. (175) the function  $\ell n(R+z)$  is harmonic except on the negative  $z$ -axis; it can be interpreted as the potential of a uniform charge distribution over that half-axis. The function  $[z \ell n(R+z) - R]$  is likewise harmonic except on the negative  $z$ -axis, and it can be interpreted as the potential of a charge distribution over that half-axis of density proportional to  $z$ .

The displacements (187) are homogeneous in  $x, y, z$  of degree  $-1$ , and give rise to strain components which are homogeneous of degree  $-2$ , resulting in stress components of the form  $f_{ij}(\theta, \phi)/R^2$  where  $R, \theta, \rho$  are spherical coordinates. Over any surface  $S$  in  $z > 0$  enclosing the origin the resultant of the surface tractions is equal to  $F_x$  in the  $x$ -direction.

Along  $z=0$ , eqs. (187) yield

$$\begin{aligned}
 u &= \frac{F_x}{4\pi G_1} \left[ \frac{2(1-\nu_1)}{R} + 2\nu_1 \frac{x^2}{R^3} \right] \\
 v &= \frac{F_y}{4\pi G_1} \left[ 2\nu_1 \frac{xy}{R^3} \right] \\
 w &= \frac{F_x(1-2\nu_1)}{4\pi G_1} \frac{x}{R^2}
 \end{aligned} \tag{191}$$

and  $z=0$  is free from tractions, except near the origin. Similarly the application at the origin of a tangential concentrated point force  $F_y$  in the  $y$ -direction gives rise to displacements that can be obtained from (187), (189), (191) by replacing  $x, y; u, v; F_x$  by  $y, -x; v, -u; F_y$ , respectively.

By applying superposition one may obtain the displacements for dis-



tributed tractions  $X(x, y)$ ,  $Y(x, y)$  over an area  $A$  in the plane  $z=0$ . Thus eq. (189) may be used for a distributed  $X(x, y)$  by replacing (188) by

$$F_1(x, y, z) = \iint_A X(x', y') \Omega dx' dy' \quad (192)$$

where

$$\Omega = [z \ln(R+z) - R], \quad R^2 = (x-x')^2 + (y-y')^2 + z^2 \quad (193)$$

Alternatively superposition may be applied directly to eq. (187) by replacing  $x, y, z$ , by  $x-x', y-y', z$ ;  $R$  by  $R'$  as in (193), replacing  $F_x$  by  $X(x', y') dx' dy$ , and integrating over  $S$ . Similar remarks apply to distributed  $Y(x, y)$ .

In particular, for  $P$  on  $z=0$  there results

$$\begin{aligned} u(P) &= \frac{1}{4\pi G_1} \left\{ \iint X(x', y') \left[ \frac{2(1-\nu_1)}{R(P, P')} + \frac{2\nu_1(x-x')^2}{R^3(P, P')} \right] dx' dy \right. \\ &\quad \left. + \iint Y(x', y') \frac{2\nu_1(x-x')(y-y')}{R^3(P, P')} dx' dy' \right\} \\ v(P) &= \frac{1}{4\pi G_1} \left\{ \iint X(x', y') \left[ \frac{2\nu_1(x-x')(y-y')}{R^3(P, P')} \right] dx' dy' \right. \\ &\quad \left. + \iint Y(x', y') \left[ \frac{2(1-\nu_1)}{R(P, P')} + \frac{2\nu_1(y-y')^2}{R^3(P, P')} \right] dx' dy' \right\} \\ w(P) &= \frac{1}{4\pi G_1} \left\{ \iint X(x', y') \frac{(1-2\nu_1)(x-x')}{R^2(P, P')} dx' dy' \right. \\ &\quad \left. + \iint Y(x', y') \frac{(1-2\nu_1)(y-y')}{R^2(P, P')} dx' dy' \right\} \end{aligned} \quad (194)$$

where presumably all the integrations are carried out over  $A$ , the area of contact. The net displacements for the upper solid are now obtained by adding (185) to (194).

For two contacting solids, initially bounded by the two surfaces as in figure 29

$$z=f_1(x, y), \quad z=f_2(x, y) \quad (195)$$

tangent to each other and to the plane  $z=0$  at the origin, eqs. (185), (194) are applied to the upper solid 1. For the lower solid 2, one replaces the elastic constants  $G_1, \nu_1$  by  $G_2, \nu_2$ , and changes the signs of the right-hand members of (194) corresponding to replacement of  $X, Y$  by their reactions  $-X, -Y$ . To avoid confusion it is advisable to introduce subscripts 1, 2 in

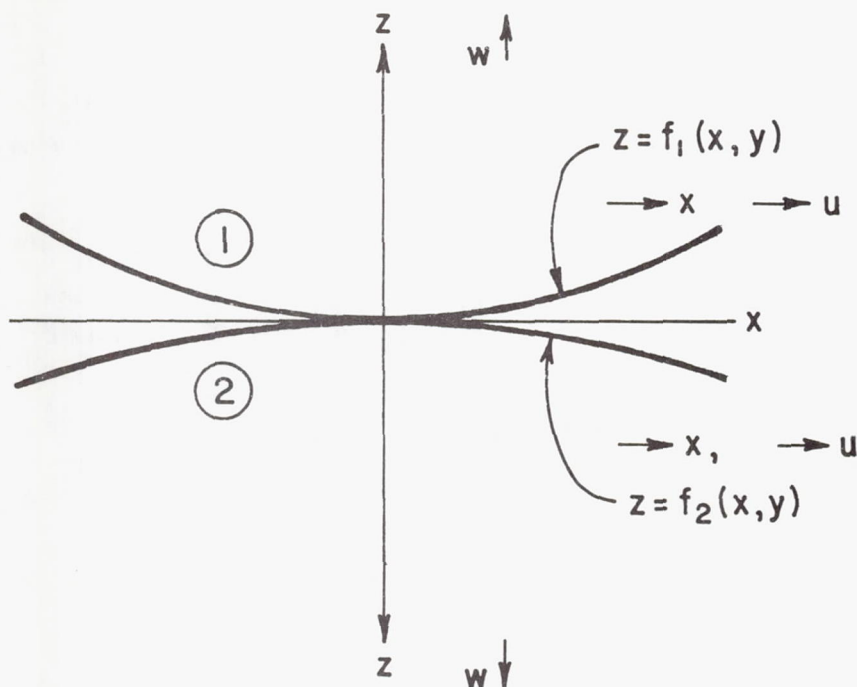


FIGURE 29.—Sign conventions for  $z$  and displacements for contacting solids.

$u, v, w$  to distinguish the displacements of the upper solid and lower solids, respectively.

The classical solution of Hertz corresponding to a compression of the solids with a compressive force  $P$  leads from (185)<sub>3</sub> to the integral equation over  $A$

$$w_1 + w_2 = C - \frac{x^2}{2R_1} - \frac{y^2}{2R_2} = \frac{1}{2\pi} \left( \frac{1-\nu_1}{G_1} + \frac{1-\nu_2}{G_2} \right) \iint_A \frac{p(P') dP'}{R(P, P')} \quad (196)$$

where  $P = \iint p(P') dP'$  and  $(x^2/2R_1 + y^2/2R_2)$  is the sum of the quadratic terms in (195), this sum being referred to its principal axes.

Hertz showed that the contact area is an ellipse

$$e: \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \quad (197)$$

and that the pressure distribution over it is ellipsoidal

$$p = \frac{2P}{3\pi ab} \left( 1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} \right)^{1/2} \quad (198)$$

He obtained this solution by utilizing the analogy (186). Hertz started with the gravitational potential  $V$  of an ellipsoid of semi-axes  $a, b, c$  with a uniform mass density  $\rho$ . Inside this potential  $V$  is quadratic in  $x, y, z$ . By letting  $c$  approach zero while  $\rho c$  is kept constant,  $V$  approaches the potential of surface distribution  $\sigma$  over the limiting ellipse (197), whose potential  $V$  is quadratic in  $x, y$  over this ellipse, and whose mass distribution is ellipsoidal, i.e., proportional to (198). Thus Hertz obtained both the shape of the contact area and of the pressure distribution over it.

Recalling eq. (194)<sub>3</sub> it will be noted that normal displacements  $w$  are also produced by the tangential tractions  $X, Y$  over the contact ellipse. The change of signs in passing from the upper to the lower solid, combined with the change of directions of positive  $z$  and  $w$ , leads to the conclusion that if the two solids are similar elastically, the contributions to  $w$  from the tangential tractions lead to the same warping of the area of contact in both contacting solids. Therefore, no correction to the shape of the area of contact, nor to the Hertzian distribution (198) is required. However, if the solids are not similar elastically, then, eq. (194)<sub>3</sub> opens a gap between solids over the contact ellipse.

Returning to eq. (185) it will be noted that, in addition to the flattening of the solids produced by  $w_1, w_2$ , there are also tangential displacements produced by the distributed pressure force  $p(x', y')$ . For the upper solid they are

$$(u_1, v_1)(P)|_{z=0} = -\frac{(1-2\nu_1)}{4\pi G_1} \iint_A p(x'y') \frac{(x-x', y-y')}{R(P, P')} dx'dy' \quad (199)$$

If the contacting solids are similar elastically, i.e., if  $G_1=G_2$ ;  $\nu_1=\nu_2$ , the displacements (199) are matched by equal displacements  $u_2, v_2$ . However, if the solids are not similar elastically, then

$$(u_2, v_2)(P)|_{z=0} = \frac{1-2\nu_2}{1-2\nu_1} \frac{G_1}{G_2} (u_1, v_1)(P)|_{z=0} \quad (200)$$

resulting in a slip over the area of contact. The tendency to slip will be resisted by friction forces, thus complicating the situation. Thus, while eq. (196) and its Hertz solution applies both to elastically similar and dissimilar solids, the tangential displacements produced by the pressure are neglected in the Hertz theory. For dissimilar solids these unequal tangential displacements result in potential slip which requires consideration of forces  $X, Y$  due to friction, even when no overall forces  $F_x, F_y$  are transmitted across the contact area.

We turn next to consideration of a rigid flat die pressed against the upper solid, and assume that the shape of the die is elliptical; in fact assume that it is given by the ellipse  $e$  in eq. (197). Now eq. (185)<sub>3</sub>,



combined with the potential analogy (186), shows that the resulting elastic problem corresponds to the potential problem of a thin, conducting, elliptic disk, charged to a constant potential  $V$ . It is known that the resulting equipotentials are ellipsoids confocal to  $e$  and that the charge density is proportional to

$$\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)^{-1/2} \quad (201)$$

Hence the pressure distribution is

$$p = \frac{P}{2\pi ab} \left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)^{-1/2} \quad (202)$$

The above analogy (or the corresponding one for the velocity potential of irrotational flow of an ideal fluid through an elliptic hole in a plane) was utilized by Mindlin (ref. 7) in what might well be termed the second classical solution of contact problems of more recent origin. It is concerned with the transmission of tangential forces  $F_x$ ,  $F_y$  across the Hertz ellipse of contact  $e$ , eq. (197). Suppose that  $F_y$  vanishes, and that the force  $F_x$  is transmitted by the lower body on the upper one, as a result of a translation  $\delta$  imposed on the solids at infinity relative to each other in the  $x$ -direction. Assume that no slipping takes place over the Hertzian ellipse of contact. If

$$\nu_1 = \nu_2 = 0 \quad (203)$$

then eq. (194) yields along  $z=0$  for the upper solid, assuming that  $Y=0$ ,

$$u_1 = \frac{2}{4\pi G_1} \iint_e \frac{X(x'y')}{R(P, P')} dx' dy' \\ v_1 = 0 \quad (204)$$

$$w_1 = \frac{1}{4\pi G_1} \iint_e \frac{X(x'y')(x-x')}{R^2(P, P')} dx' dy'$$

while for the lower one there results

$$u_2 = -\frac{2}{4\pi G_2} \iint_e \frac{X(x', y')}{R(P, P')} dx' dy' \\ v_2 = 0 \quad (205)$$

$$w_2 = \frac{1}{4\pi G_2} \iint_e \frac{X(x', y')(x-x')}{R^2(P, P')} dx' dy'$$

Similar to (173), (187), the displacements (187) are homogeneous in  $x$ ,  $y$ ,  $z$  of degree  $-1$  and vanish at infinity; likewise the displacements (194), (204), (205) vanish at infinity. We therefore add to (204), (205)

a translation of the upper solid in the  $x$ -direction, of amount  $-\delta_1$  and of the lower of amount  $\delta_2$ , such that

$$\delta_1 + \delta_2 = \delta, \quad \delta_1 G_1 = \delta_2 G_2 \quad (206)$$

Then the conditions of no net displacement over the ellipse  $e$  given by (197) for the upper solid yield

$$u_1 G_1 = \delta_1 G_1 + \frac{1}{2\pi} \iint_e \frac{X(x', y')}{R(P, P')} dx' dy' = 0 \quad (207)$$

where

$$\iint_e X(x' y') dx' dy' = F_x$$

and essentially the same integral equation results for the lower solid. It will now be seen that the problem of finding the distribution of  $X$  is equivalent to the problem of finding the charge distribution over a thin, conducting, elliptically-shaped disk, charged to a constant potential; or again to the pressure distribution over a flat, rigid, elliptically-shaped, die considered above. Hence the distribution of  $X$  is given by

$$X(x, y) = \frac{F_x}{2\pi ab} \left( 1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} \right)^{-1/2} \quad (208)$$

Mindlin also found the solution for non-vanishing Poisson's ratio for elastically similar solids. He arrives at this solution in two different ways, one is by using the Boussinesq harmonic functions  $L$ ,  $M$ , related to the displacements  $u$ ,  $v$  on the boundary  $z=0$ , thus

$$L = \iint_{-\infty}^{+\infty} \frac{u(x', y')}{R(P, P')} dx' dy', \quad M = \iint_{-\infty}^{+\infty} \frac{v(x', y')}{R(P, P')} dx' dy' \quad (209)$$

where the integration is carried out over the complete plane  $z=0$ ; in terms of  $L$ ,  $M$  the displacements  $u$ ,  $v$  in  $z>0$  are given by

$$2\pi u = -\frac{\partial L}{\partial z} + \frac{z}{2} \frac{\partial}{\partial x} \left( \frac{\partial L}{\partial x} + \frac{\partial M}{\partial y} \right) \quad (210)$$

$$2\pi v = -\frac{\partial M}{\partial z} + \frac{z}{2} \frac{\partial}{\partial y} \left( \frac{\partial L}{\partial x} + \frac{\partial M}{\partial y} \right)$$

The other way is by means of the harmonic stress function  $F_1$  of eqs. (188), (192). Mindlin shows that eq. (208) still holds and that  $Y$  vanishes on  $z=0$ , as well as the displacement component  $v$  on  $e$ .

Mindlin also solved the torsional problem where a moment  $M_z$  is

transmitted across the Hertzian contact ellipse, again for the condition of no slip. He obtained for the tractions

$$X = T'y \left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)^{-1/2}, \quad Y = -T''x \left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)^{-1/2} \quad (211)$$

where  $T'$ ,  $T''$  are proper constants. When the ellipse  $e$  reduces to the circle  $r=a$ , the traction is azimuthal and is given by

$$Y = \frac{3M_z}{\pi a^4} y \left(1 - \frac{r^2}{a^2}\right)^{-1/2}, \quad Y = -3M_z x \left(1 - \frac{r^2}{a^2}\right)^{-1/2} \quad (212)$$

The tangential tractions (208), (211) become infinite at the edge of the contact area, where the pressure  $p$  vanishes. Hence slip must occur near the edge, past the locus where  $X = \mu p$ . If complete slip occurs, then  $X \equiv \mu P$ , and  $F_x = \mu P$ . For the circular contact area  $X$  is then given by eq. (150). By evaluating the function  $F_1$  Mindlin obtained the displacements  $u$ ,  $v$  for  $F_x < \mu P$ , and he arrived (independently of Cattaneo) at the solution (150) to (153). This solution corresponds to no slip over the circle  $r < a'$  and shows that a constant slip in the  $x$ -direction has taken place over the area  $a' < r < a$ .

The above results apply to stationary contacts. If the tractive force  $F_x$  (and/or the normal force  $P$ ) varies with time, the load displacement curves exhibit hysteresis (Mindlin and Deresiewicz, ref. 31; Mindlin, et al., ref. 32).

Rolling contact between two spheres with creep and spin have been considered by de Pater (ref. 5) and Kalker (ref. 33). Recall eq. (59). Elimination of  $\xi$  or  $\eta$  leads to

$$\frac{x^2}{a^2 \cosh^2 \eta} + \frac{y^2}{a^2 \sinh^2 \eta} = 1, \quad \frac{x^2}{a^2 \sin^2 \xi} - \frac{y^2}{a^2 \cos^2 \xi} = 1 \quad (213)$$

which represent, respectively, ellipses and hyperbolas with foci at  $x = \pm 1$ ,  $y = 0$ . Carrying out the replacements

$$x \rightarrow r, \quad y \rightarrow z, \quad \cos \xi \rightarrow \omega, \quad \sinh \eta \rightarrow \eta \quad (214)$$

where  $r$  is the distance from the  $z$ -axis, eq. (213) leads to a system of oblate spheroids and normal hyperboloids of one sheet, both axially symmetric about the  $z$ -axis. Eq. (59) now yields

$$z = a\omega\eta, \quad r = a\sqrt{(1-\omega^2)(1+\eta^2)} \quad (215)$$

Axially symmetric harmonic functions of the product form are given by

$$P_n(\omega)Q_n(i\eta), \quad n = 0, 1, \dots \quad (216)$$

where  $P_n(\xi)$  is the Legendre polynomial of degree  $n$ ,  $Q_n(\xi)$  the Legendre function of the second kind, which is a second solution of the second-order



differential equation satisfied by  $P_n(\xi)$ . This is evaluated for the pure imaginary argument  $i\eta$ .

With the azimuthal angle  $\psi$  added as a third coordinate, product harmonic functions are given by

$$\frac{\cos}{\sin} m\psi P_n^m(\omega) Q_n^m(i\eta) \quad (217)$$

where  $P_n^m$  are the associated Legendre polynomials of degree  $n$  and order  $m$ ,  $Q_n^m$  the corresponding functions of the second kind, defined by

$$P_n^m(\xi) = (1 - \xi^2)^{m/2} \frac{d^m P_n(\xi)}{d\xi^m} \quad (218)$$

$$Q_n^m(\xi) = (1 - \xi^2)^{m/2} \frac{d^m Q_n(\xi)}{d\xi^m}$$

Again, in (217) the functions of the second kind are evaluated for complex values  $i\eta$  of their argument.

de Pater and Kalker assume that both the creep and the spin rate are small, and that the slip area is confined to a thin strip at the exit side of the circle of contact. The creep conditions over  $L$  (the circle of contact  $r < a$ ) now become

$$\frac{\partial u}{\partial x} = -\frac{1}{2}s_x + \frac{1}{2}\phi y, \quad \frac{\partial v}{\partial x} = -\frac{1}{2}s_y - \frac{1}{2}\phi x \quad (219)$$

where  $s_x$ ,  $s_y$ ,  $\phi$  are the components of creep and the spin. Integrating with respect to  $x$  yields

$$\begin{aligned} u &= -\frac{1}{2}s_x x + \frac{1}{2}\phi xy + f(y/a) \\ v &= -\frac{1}{2}s_y x - \frac{1}{4}\phi x^2 + g(y/a) \end{aligned} \quad (220)$$

where  $f$ ,  $g$  are arbitrary functions of their argument. Over the forward edge of the contact area, by analogy to the two-dimensional case, it is assumed that

$$(X, Y) \sim \sqrt{a^2 - r^2} \quad \text{for } r \rightarrow a, x > 0 \quad (221)$$

Over the exit side of the contact circle boundary, the tractions become infinite as

$$(X, Y) \sim \frac{1}{\sqrt{a^2 - r^2}} \quad \text{for } r \rightarrow a, x < 0 \quad (222)$$

The functions  $L$ ,  $M$  were used by de Pater. He treated the case  $\nu = 0$ , chose for  $f$  and  $g$  quadratics in  $y^2$ , but satisfied the boundary conditions

(221) only in a very approximate way. Kalker used not the Boussinesq functions but the so-called Papkovitch-Neuber harmonic functions (refs. 34 and 35), and made no assumptions regarding the value of  $\nu$  nor the shape of the functions  $f, g$ . He expanded his harmonic functions in series of the product harmonics (216), (217) in which he replaced  $P_n^m$  by  $P_n^{-m}$ , the associated Legendre functions of negative order. After applying (221) he determined  $f, g$ , and by a very complex analysis arrived at an infinite number of linear equations in the coefficients of the expansions. These equations were reduced to a finite number (eight) by truncation and solved numerically for various values of  $\nu$ . As a check on the accuracy, the case  $\nu=0$  (for which Kalker obtained a finite form solution by a different method) was used, and it showed good agreement of the coefficients from the truncated equations with their exact values.

Comparison of the results of Kalker's calculations with Johnson's calculations in reference 15, eq. (166), and his experiments, showed that Johnson's coefficients in the linear expressions of  $F_x, F_y, M_z$  in terms of the creep components and the spin were generally higher than Kalker's, except for the coefficient of the spin in  $M_z$  where the Kalker value is in better agreement with Johnson's experimental results.

We close this section devoted to analytical solutions with a review of a paper by Mow, Chow, and Ling (ref. 8), in which they consider the Hertz-Mindlin problem for contacts between solids bounded by shapes involving both quadratic and fourth-degree terms in  $x, y$ , and subjected to normal and shearing forces over the contact area which is still assumed to be elliptical in shape. They consider stationary (i.e., non-rolling) contacts, a finite coefficient of friction, and are led to a central locked area that is also elliptical and similar to the contact ellipse. This is accomplished by proper choice of the shape of the pressure and shear over the contact ellipse. First they show that if a pressure distribution of the form

$$p(x, y) = (L')^{1/2} \left( \alpha_1 + \alpha_2 \frac{x^2}{a^2} + \alpha_3 \frac{y^2}{b^2} \right) \quad (223)$$

where

$$L' = \left( 1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} \right) \quad (224)$$

is assumed over the ellipse (197), then  $w$  can be determined, over (197), and it turns out to be a polynomial in  $x^2, y^2$

$$w = b_0 + b_1 x^2 + b_2 y^2 + b_3 x^2 y^2 + b_4 x^4 + b_5 y^4 \quad (225)$$

whence the initial shape (difference) of the contacting solids over (197) can be determined. Next, to handle the Mindlin problem, (223) is specialized to

$$p(x, y) = (L')^{1/2}(\alpha_1 + \alpha_2 L') \quad (226)$$

to obtain for the locked area  $L$  an ellipse of semi-axes  $\xi_a, \xi_b, \xi < 1$ , namely, the interior of the ellipse

$$E': 1 - \left(\frac{x}{\xi a}\right)^2 - \left(\frac{y}{\xi b}\right)^2 = 0 \quad (227)$$

The shear distribution is chosen as follows

$$X = \begin{cases} \mu p(x, y) - (L')^{1/2}[\beta_1 + \beta_2 L'] & \text{over } L \\ \mu p(x, y) & \text{over } S \\ 0 & \text{outside } e \end{cases} \quad (228)$$

Then  $u$  turns out to be constant over  $L$ , where  $X < \mu p$ , while slip occurs over  $S$ .

The solution is carried out by introducing general confocal coordinates, connected with the family of quadric surfaces

$$\Theta = \frac{x^2}{a^2 + \theta} + \frac{y^2}{b^2 + \theta} + \frac{z^2}{\theta} - 1 = 0 \quad (229)$$

Eq. (229), for each fixed  $\theta$ , represents,

- (1) An ellipsoid, if  $0 < \theta < \infty$
- (2) A hyperboloid of one sheet, if  $-b^2 < \theta < 0$
- (3) A hyperboloid of two sheets, if  $-a^2 < \theta < -b^2$

For a given point  $P: (x, y, z)$  the confocal coordinates are defined as the values of  $\theta$  for which the surface (229) passes through  $P$ . It turns out that for a general position of  $P$  there are three such values,  $\lambda, \mu, \nu$ , located in the intervals (230) in that order, and that the three surfaces (229) are mutually perpendicular and include  $E$ , the ellipse (197) as a limiting case as  $\lambda$  approaches 0 through positive values.

Turning to harmonic functions of the product form

$$V = \Lambda(\lambda) M(\mu) N(\nu) \quad (231)$$

there are such functions that are polynomials in  $x, y, z$  of degree  $n$ , and they are called ellipsoidal harmonics of degree  $n$ . In addition to the form (231) which will be discussed presently, they can also be put in the form

$$V = \begin{Bmatrix} x & yz \\ 1 & y & zx & xyz \\ z & xy \end{Bmatrix} \Theta_1 \Theta_2 \dots \Theta_m \quad (232)$$



where  $\Theta_i$  is the middle term of (229), evaluated for a proper, constant value  $\theta = \theta_i$  of  $\theta$ . The preceding factor may be chosen as any one of the eight terms in the brace, giving rise to ellipsoidal harmonics of the first, second, third, and fourth species, depending upon the column from which this factor is chosen. It will be seen that if  $n$  is even,  $V$  will be of the first or third species and  $m = n/2$ ,  $(n-2)/2$ ; if  $n$  is odd,  $V$  is of the second or fourth species, and  $m = (n-1)/2$ ,  $(n-3)/2$ . It will be noted that these harmonics change sign over particular quadrics of the confocal family (229), as well as over some of the coordinate planes.

A third form for the ellipsoidal harmonics of degree  $n$  is given by

$$V = \left( 1 - \frac{D^2}{2(2n-1)} + \frac{D^4}{2.4(2n-1)(2n-3)} + \cdots \right) H_n \quad (233)$$

where  $H_n$  is a harmonic polynomial, homogeneous of degree  $n$ , and

$$D^2 = a^2 \frac{\partial^2}{\partial x^2} + b^2 \frac{\partial^2}{\partial y^2} \quad (234)$$

is the Niven operator. The functions  $H_n$  are well known, a complete set of them,  $2n+1$  in number, being given by  $R^n P_n^m(\cos \theta)(\cos m\psi, \sin m\psi)$ , where  $R$ ,  $\theta$ ,  $\psi$  are spherical coordinates. Hence the total number of ellipsoidal harmonics of degree  $n$  is, likewise,  $2n+1$ .

Return now to the form (231). Separating variables in the Laplace equation, it turns out that for functions of the first species  $\Lambda$  is a polynomial of degree  $m = n/2$  in  $\lambda$ :

$$\Lambda(\lambda) = P_m(\lambda) \quad (235)$$

and that it is a solution of the differential equation

$$4\Delta_\lambda \frac{d}{d\lambda} \left( \Delta_\lambda \frac{dP}{d\lambda} \right) = [n(n+1)\lambda + C]P \quad (236)$$

where

$$\Delta_\lambda = [(a^2 + \lambda)(b^2 + \lambda)\lambda]^{1/2} \quad (237)$$

and  $C$  is a proper constant. Moreover  $M(\mu)$  and  $N(\nu)$  are exactly the same functions of  $\mu$  and  $\nu$  that  $\Lambda$  is of  $\lambda$ .

For functions of the second, third, and fourth species,  $\Lambda(\lambda)$  turns out to be of the form

$$\Lambda(\lambda) = (\sqrt{a^2 + \lambda}, \sqrt{b^2 + \lambda}, \sqrt{\lambda}) P_m(\lambda) \left( m = \frac{n-1}{2}, \frac{n-2}{2}, \frac{n-3}{2} \right) \quad (238)$$

where  $P_m$  is a polynomial in  $\lambda$  of the respective degree  $m$  indicated, and one, two or three factors of the preceding parenthesis are used, and  $P_m$  satisfies the differential equation of the form (236) with  $n$  replaced by  $n-1$ ,  $n-2$ ,  $n-3$ , respectively (and proper  $C$ ). Again  $M$ ,  $N$  are the same

functions of  $\mu$  and  $\nu$  as  $\Lambda$  is of  $\lambda$ . The  $m$  roots of  $P_m$  are all real and distinct, and they are precisely equal to the values  $\theta_i$  of  $\theta$  in the factors  $\Theta_i$  in the form (232).

Equation (236) is known as the Lamé differential equation. An alternative form for it is obtained by introducing the variable  $u$  by means of

$$u = \int^{\lambda} \frac{d\lambda}{\Delta_{\Lambda}} = \int^{\lambda} \frac{d\lambda}{[4(a^2 + \lambda)(b^2 + \lambda)\lambda]^{1/2}} \quad (239)$$

This integral is elliptic; it can be put in the standard Weierstrass form

$$u = \int_{\xi}^{\infty} \frac{d\xi}{(4\xi^3 - g_2\xi - g_1)^{1/2}} = p^{-1}\xi \quad (240)$$

by letting

$$\xi = \lambda + (a^2 + b^2)/3 \quad (241)$$

leading to

$$\xi = p(u) = \lambda + (a^2 + b^2)/3 \quad (242)$$

Here  $p(u)$  is the Weierstrass "p-function" (generally denoted by  $\xi$ ); it is doubly-periodic in the complex  $u$ -plane;  $p^{-1}$  is its inverse function.

With  $u$  as a basic variable, the Lamé equation now becomes

$$\frac{d^2\Lambda}{du^2} = [n(n+1)p(u) + B]\Lambda = 0 \quad (243)$$

where  $B$  is a (proper) constant. In this form it is satisfied by the functions  $\Lambda$  in (235), (238) of all four species. The solutions in question are called Lamé functions of the first kind and proper species, and are denoted by  $E_n^m$  where  $n$  is the degree, and  $m$  an index referring to the values of the constant  $B$  for which these solutions exist.

The ellipsoidal harmonics (231) are formed by multiplying the functions  $E_n^m$  of  $\lambda$  or  $u$  by corresponding functions of  $\mu$ ,  $\nu$ , or of  $v$ ,  $w$  obtained by replacing  $\lambda$  by  $\mu$ ,  $\nu$ ;  $u$  by  $v$ ,  $w$ , in eqs. (239) to (243).

For contact problems over the ellipse  $E$ , the ellipsoidal harmonics will not do, even though they reduce to polynomials in  $x$ ,  $y$  over  $E$ , because they become infinite at infinity, while the elastic displacements and stresses vanish at infinity. To remedy this situation one retains the Lamé functions of the first kind in  $\mu$  and  $\nu$ , but replaces  $E_n^m(\lambda)$  by a different solution of Lamé equation, namely, one that vanishes at infinity. This solution, denoted by  $F_n^m$ , can be expressed in terms of  $E_n^m$  by quadratures, and is defined as

$$F_n^m(\xi) = (2n+1)E_n^m(\xi) \int_0^u \frac{du}{[E_n^m(\xi)]^2} \quad (244)$$

The effect of introducing  $F_n^m$  is thus to multiply the ellipsoidal harmonic solution by the integral in (244).

Throughout the above we have used the Weierstrass form (ref. 36), as a result of which the same Lamé differential equation was obtained for all three factors in (231), (243). While the values of  $u$  are real, the values of  $v$  and  $w$  are not. In applications the Jacobian form of the elliptic functions and integrals is used by the authors. Thus  $\lambda$  is replaced by the authors by  $\rho$  where

$$\lambda = a^2(1 - \rho^2) \quad (245)$$

whereupon

$$\frac{d\lambda}{\Delta_\lambda} = \frac{d\rho}{[(1 - \rho^2)(e^2 - \rho^2)]^{1/2}} \quad (246)$$

where  $\bar{e}$  is the eccentricity of the ellipse  $e$ :  $\bar{e} = (1 - b^2/a^2)^{1/2}$ . The main objective of the authors is to show that by proper choice of the constants  $\alpha_1, \alpha_2$  in (226), they can effect an appreciable reduction in slippage over the area  $S$  as compared to the Mindlin solution for the same  $P, F_x$ , and  $\mu$ .

For smooth surfaces, parabolic approximations to the shapes of the contacting solids are generally sufficient in view of the small size of the area of contact. It is therefore presumed that the results of this paper are primarily of interest for cases where extensive contacts are secured by having a high degree of conformity between the contacting solids.

## 7. NUMERICAL METHODS

With the relative lack of success for theoretical solutions in rolling contacts with creep and spin, and the availability of high-speed computers, it is not surprising that direct numerical methods of solving these contact problems have been tried but thus far with only moderate success. Kalker (ref. 37) uses a method based on expressing the elasticity solution in  $z \geq 0$  in terms of ellipsoidal harmonics, similar to those used in the sixth section, and for which  $X, Y$  over  $e$  are equal to the product of the Hertzian pressure by polynomials in  $x, y$ , while the displacements  $u, v$  are likewise polynomials  $x, y$  over  $e$ . We write these solutions over  $E$  in the form

$$X = ZX_i, \quad Y = ZX_i, \quad u = u_i, \quad v = v_i \quad (247)$$

where  $Z$  is the Hertzian pressure (198) and  $X_i, Y_i; u_i, v_i$  the polynomials in  $x, y$  characteristic of these solutions. For the calculation of the latter, Kalker refers to Galin (ref. 38) and Dornorovich (ref. 39). When  $e$  reduces to a circle of contact, Kalker refers to his own thesis (ref. 33). Using a finite linear combination of these solutions and denoting the coefficients by  $\mu C_i$  the traction conditions of slip and no-slip over  $S$  and  $L$  become

$$(X^*)^2 + (Y^*)^2 - 1 < 0 \quad \text{over } L \quad (248)$$

$$(X^*)^2 + (Y^*)^2 - 1 = 0 \quad \text{over } S \quad (249)$$

where



$$X^* = \sum C_i X_i, \quad Y^* = \sum C_i Y_i \quad (250)$$

Denote the slip components by  $s_x, s_y$ ; their resultant by  $s$ ; and let  $w_x, w_y$  be a unit vector in the direction of the slip

$$w_x = s_x/s, \quad w_y = s_y/s \quad (251)$$

Then the slip and no-slip conditions also imply

$$s_x = 0, \quad s_y = 0 \quad \text{over } L \quad (252)$$

$$w_x = X^*, \quad w_y = Y^* \quad \text{over } S \quad (253)$$

They can be combined into

$$S = s_x^2 + s_y^2 = 0 \quad \text{over } L \quad (254)$$

$$T = (x_x - X^*)^2 + (w_y - Y^*)^2 = 0 \quad \text{over } S \quad (255)$$

The explicit expressions for  $s_x, s_y$  are

$$s_x = \xi_x - \frac{\phi y}{U} + \frac{\partial u}{\partial x}, \quad s_y = \xi_y + \frac{\phi x}{U} + \frac{\partial v}{\partial y} \quad (256)$$

and replacing  $u, v$  by the proper combination of the solutions (247)

$$s_x = \xi_x - \frac{\phi y}{U} + \mu \sum C_i \frac{\partial u_i}{\partial x}$$

$$s_y = \xi_y + \frac{\phi x}{U} + \mu \sum C_i \frac{\partial v_i}{\partial x}$$

Kalker combines the separate conditions over  $L$  and  $S$  into

$$TS = 0, \quad (X^*)^2 + (Y^*)^2 \leq 1 \quad \text{over } e \quad (257)$$

Since the finite series solution can only fulfill the above conditions approximately, while  $T$  and  $S$  will always be positive, he replaces the above by

$$I = \iint_e TS \, dx' dy' = \text{Minimum} \quad (258)$$

$$(X^*)^2 + (Y^*)^2 \leq 1 \quad \text{over } e$$

The number of coefficients  $C_i$  was  $(m+1)(m+2)$  where  $m$  is the highest degree of the  $(X_i, Y_i)$ -polynomials employed. The calculations of the optimum were carried out by solving the equations

$$\frac{\partial I}{\partial C_i} = 0 \quad (259)$$

Since  $C_i$  enters into the integrand through  $w_x, w_y$  which are not linear in

the coefficients, the solution of (259) was not easy. It was carried out by iteration, by starting with a set of coefficients and improving them by Newton's method, by solving the linear equations in  $\Delta C_i = (C_i)_{n+1} - (C_i)_n$

$$0 = \left( \frac{\partial I}{\partial C_i} \right)_{n+1} = \left( \frac{\partial I}{\partial C_i} \right)_n + \sum_i \left( \frac{\partial^2 I}{\partial C_i \partial C_i} \right)_n \Delta C_i \quad (260)$$

where the subscript  $n$  refers to  $n$ -stage of the iterations,  $n+1$  to the next stage; the differentiations were carried through the integral sign and applied to the integrand  $TS$ , It is not indicated how the differentiations were carried out, but from the fact that over one minute of the computer time was required to set up the coefficients of the linear equations (260) (80 in number for  $m=4$ ) and solve them, it would appear that the differentiations were carried out analytically. The integrations were carried out by summing over a mesh of points over the contact area; 100 points were used for  $m=4$ .

In a series of calculations in which one of the parameters  $\xi_x$ ,  $\xi_y$ ,  $\phi$  was changed by a small amount, the first set of assumed values of  $C_i$  coefficients was chosen as the true (i.e., converged) values of the preceding solution, and the number of iterations was not very large. While the method is applicable to elliptic contact areas, all the examples cited are for a circular contact area, where the ellipsoidal harmonics simplify, for  $\eta=0$ , to  $P_n^m(\omega) (\sin m\psi, \cos m\psi)$ , which are indeed polynomials in  $x, y$ .

The inequality in (258) was generally disregarded in the calculations, its violation over  $S$  implying a local departure of the coefficient of friction from constancy, where the violation of (255) indicates an angular spread between the friction force and the traction. As the curve  $\Gamma$  separating the locked from the slip regions never enters in the calculations, the determination of these areas was carried out by assigning all the mesh points with  $S$  exceeding  $T$  to the slip area. Those where  $T$  exceeded  $S$  were assigned to the locked area, provided the differences  $|S - T|$  were large. The curve  $\Gamma$  was then obtained by (graphical) interpolation between the two areas.

The results of the calculations for single  $\xi_x$ ,  $\xi_y$ ,  $\phi$  were compared with the experimental data for the forces  $F_x$ ,  $F_y$ , showing good agreement (with properly estimated  $\mu$ ,  $\nu$ ). The shape of the curve  $\Gamma$ , the stresses  $X$ ,  $Y$ , and the slip for  $y=0$  were compared with the photoelastic results obtained by Haines and Ollerton (ref. 26) for the case of non-vanishing  $s_x$  (and  $s_y=0$ ,  $\phi=0$ ), showing fair agreement. For spin only, especially for large spin, there were indications of three small  $L$ -areas, two of which the author regards as spurious in view of the large angles between slip and traction and  $Y$  exceeding  $\mu p$  along  $y=0$ .

Quite a different numerical method was tried by Nayak and Paul (ref. 40). The authors note that over the slip area  $S$  the resultant tan-

gential traction is known, being equal to  $\mu p$ . For creep in the rolling direction or transversely to it (and no spin), the traction may be assumed to be in the direction of the creep. In these cases  $X$  is thus known over  $S$  while  $Y$  vanishes there, or the converse is true. It is only necessary to determine the traction distribution over the locked area  $L$ .

To this end the authors resort to eqs. (194) expressing the displacements  $u, v$  over the contact area in terms of the tractions over  $e$ . The contribution to these integrals arising from  $S$  can be completely determined since  $X, Y$  are known there; the contributions from  $L$  are left as integrals over that area.

Next the zero slip conditions are applied over  $L$ , say in the form of the pertinent eq. (256) with the left-hand member annihilated. The differentiations of  $u$  are presumably carried out by differentiating in (195) under the integral signs. In this way one has reduced the problem to the solution of a single linear integral equation in the traction over the locked area once the curve  $\Gamma$  has been assumed. A most important condition is that over  $\Gamma$  the  $L$ -traction should join on continuously the  $S$ -traction.

The method was tried out for the case of longitudinal creep and a circular contact area. The integrals were replaced by finite sums over a mesh of 50 points. Since the integrals are singular when the point of integration  $P'$  coincides with the field point, the contribution to the integral from the traction at that point was computed by interpolating the values of the integrand from surrounding points—a rather severe way of smoothing out the singularity of the integrand.

The determination of the shape of  $\Gamma$  was attempted by assuming for it an equation of the form

$$y = \sum_{n=0}^N \alpha_n x^n \quad (261)$$

where the coefficients were initially chosen so as to yield a shape agreeing with experimental shapes. Then the coefficients were perturbed so as to minimize the integral over  $\Gamma$  of the square of the difference between the values of  $X$  on the two sides of  $\Gamma$ . This procedure, however, failed to converge to a limiting shape for  $\Gamma$ .

The calculation did yield very good agreement with Johnson's experimental results so far as the total magnitude of the force  $F_x$  is concerned. However, the traction distribution over  $y=0$  showed no agreement with the results of Haines and Ollerton even for their shape of the curve  $C$ .

Recalling the successful use of the method of integral equations by Johnson for contact of cylinders with different elastic constants, it will be noted that the kernel (integrand multiplying the unknown function) was singular there too, though it became infinite in a much milder manner, namely, logarithmically. Johnson's linear equations were obtained by



assuming a shape for  $p$  and  $q$  (a linear variation over the interval was used) and carrying out the integration over each interval analytically. The kernels in (194) are integrable at their singular points; their derivatives, however, are not integrable, so that their integrals over an area excluding the neighborhood of the singularity by means of a proper closed curve will approach a different limit or will diverge as the curve shrinks down to the singular point depending upon its shape.

Yet the method of integral equations has been used successfully in solving many potential boundary value problems that are practically insoluble by classical methods. One such problem, the determination of the total charge on a conducting cube charged to a constant potential, is solved by Reitan and Higgins (ref. 41). This is done by breaking up the surface of the cube into a finite number of equal squares, assuming a constant charge density over each small square (the potential of this distribution over a square can be integrated), setting up linear equations for the charge densities by equating the potential at the square centers to the constant potential and solving these equations. Axially symmetric potential fields outside a cylindrical boundary were similarly solved by Mei and Van Bladel (ref. 42) and Van Bladel (ref. 43).

There is no doubt that once the singularities of the kernels in (194) are tamed in a similar fashion, the method of integral equations can be used to advantage in solving contact problems with creep and spin. Moreover, judging by the results obtained by Johnson, this method will probably be the first one to elucidate the rolling contact phenomena of elastically dissimilar solids.

#### 8. CONCLUDING REMARKS

It has been pointed out by Nayak and Paul (ref. 40) that when the theoretical results of the creep and slip are compared with experimental measurements, wide discrepancies are observed. Figure 30 indicates the comparison in question, between two variables proportional respectively to  $F_x$ ,  $\xi_x$  for rolling contact over a circular contact area, for six sets of experimental results, and their comparison with the theory. While each set of data lines up close to a curve, wide variation is observed between these curves, and the calculations are based on existing theory. The authors believe that these discrepancies are due to the assumption of an ideally smooth surface and of a constant coefficient of friction. Actually the surfaces depart from the ideally smooth shapes, and the size of the asperities and their statistical distribution in height must be taken into account, both in modifying the Hertz theory of contact and in the study of slip. Presumably contact actually occurs between the high spots of each surface; the friction force is pictured as arising from the plastic yielding at the high spots accompanied by welding and tearing. The

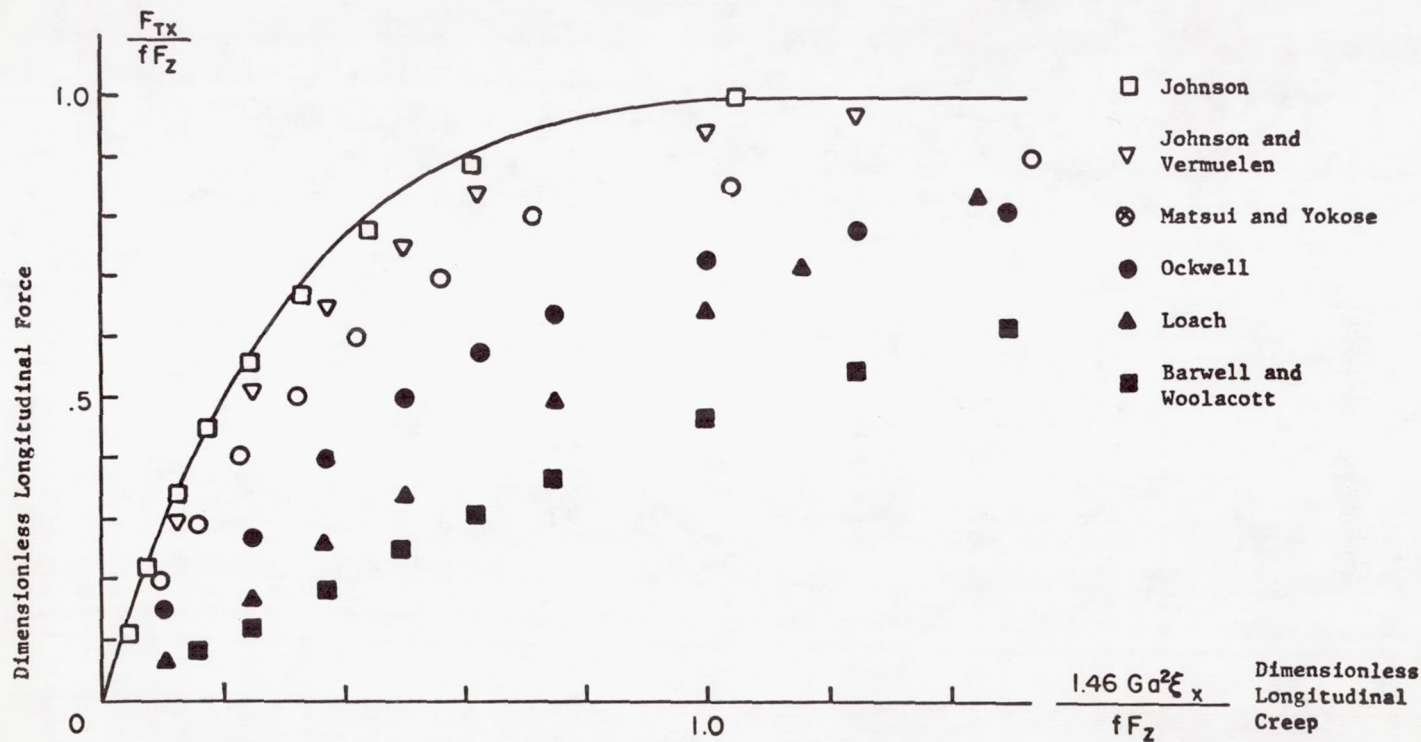


FIGURE 30.—Experimental results for longitudinal creep. The solid line is the prediction of the smooth-surface theory (ref. 40).

authors propose a rough surface theory and arrive at different relations between the variables of figure 30 which depend upon a further parameter involving the average size of the junction spots.

The title of this symposium refers to the lubrication of concentrated contacts. Yet, thus far no mention has been made of the effect of lubricants in the discussion of slip and creep, and indeed, it is not clear whether creep and slip phenomena persist under conditions of elastohydrodynamic lubrication (EHL), or whether the shearing of the lubricant renders the previous theory completely inapplicable. In reference 14 Johnson had to assume the value  $\mu=0.11$  for the coefficient of friction for dry contacts and  $\mu=0.09$  for lubricated contacts to obtain agreement with his calculations for low spin. More generally the effects of lubrication in decreasing friction and prolonging the life of rolling contacts are well known. The theory of EHL is well established by now; for high loads it leads to an essentially Hertzian pressure distribution over most of the contact area with a nearly constant film thickness. Yet the computed, nearly constant thickness of the film at high loads is often of the same magnitude as the rms of the height of the surface asperities, suggesting boundary lubrication over the high spot contacts with hydrodynamic lubrication in between. Most of the EHL studies do not take into account the effect of the tangential displacement of the bearing surfaces (nor of the modification of the normal displacements by the pressure forces when the contacting solids are dissimilar elastically).

A notable exception is the study by Johnson and Cameron (ref. 44) of tractions transmitted by EHL oil films in a rolling disc machine under conditions of contact pressures up to 260 000 psi. Figure 31 shows the variation of  $F_x/P$  with the sliding velocity and with rolling speed. The effect of the variation of the oil viscosity with the temperature of the oil is taken into account. At high sliding speeds,  $F_x/P$  approaches a ceiling, shown as the topmost curve on the left half of figure 31. The ceiling appears to be independent of contact pressure, rolling speed, and disc temperature, but depends only on the sliding speed. The authors conclude that this ceiling cannot be explained by the reduction in viscosity due to the heating up of the oil film by frictional heating. A possible explanation might be based on a plastic shear of the oil past a certain critical shear stress.

An experimental and analytical study of EHL films in rolling contact of rough surfaces is described by Tallian, et al. (ref. 45). They utilized a rolling four-ball configuration with a radiotracer tagged ball; wear rates were determined in the microgram range. The wear rate is greatest at low spindle speeds and increases with Hertzian load where no EHL film is formed. Existence and interruptions of the EHL films were observed by measuring the electrical conductivity through the Hertzian contact. The conductivity results were correlated with the expected interruptions



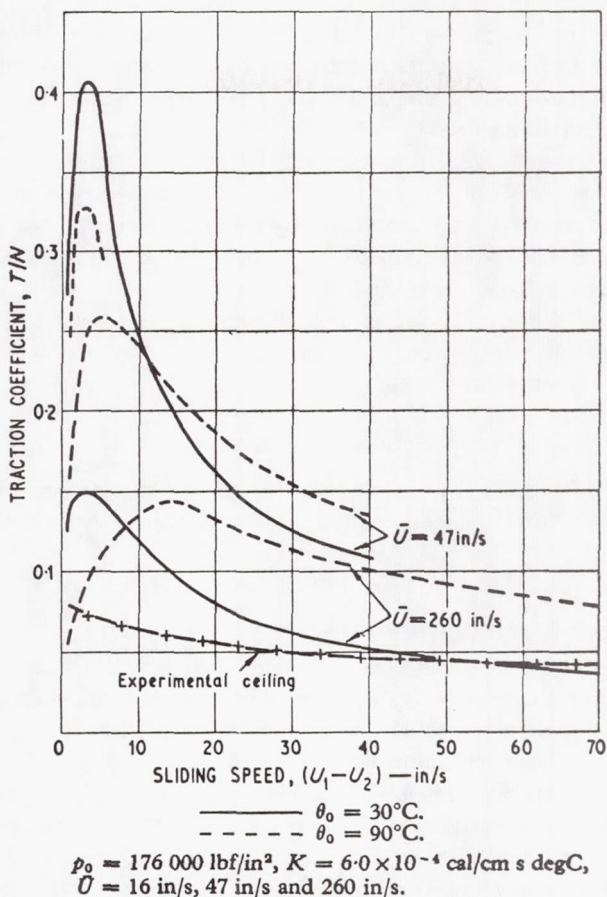


FIGURE 31.—Variation of traction coefficient with sliding speed calculated by Crook's theory at various rolling speeds and disk temperatures (ref. 44).

in the voltage across the film when the asperities break the film by making contact.

Due to limitations of space and time the discussion of 3-dimensional contacts has not been as thorough as that of cylindrical contacts. It is evident that the theory of the former is much less developed than for the latter. In particular, the interaction of the normal and tangential tractions in affecting the contact areas deserves a more thorough examination, both for elastically similar solids and especially for dissimilar ones, where changes in the Hertzian pressure distribution and even in the elliptic shape of the contact area are to be expected. Most likely, judging by Johnson and Bentall's results, the first answer to these questions may come

from numerical solutions, but, as pointed out, these must wait for the development of proper approximations to the singular integral equations which are now more "singular" and involve three (rather than two) unknown distributions.

Many of the bearing applications involve high conformity contacts. These are secured by the rolling of a ball on a highly conforming race whose curvatures are in the same direction as those of the ball. Under these conditions, the spreading of the load from the contact area may be quite different from its spreading in a semi-infinite space. Indeed, the angle subtended by the major axis of the ellipse of contact at the ball center is often quite large. An exact solution, even for a cylinder over diametrically opposite parts of the boundary of a slightly elliptical hole would be of interest.

#### ACKNOWLEDGMENTS

Figures 17, 18, 21, 22, 24, 25, 26, 27, 28, 30, and 31 have been reproduced from the referenced papers by permission of the respective publishers.

#### DISCUSSION

##### A. Seireg (University of Wisconsin, Madison, Wisconsin)

Dr. Poritsky has presented a comprehensive and valuable review of microslip in rolling contacts. One aspect of microslip that may be relevant to this comprehensive review is Mindlin's model for the hysteretic behavior of Hertzian contacts within the region of no gross slip (refs. 7, 32, and 46). Another phenomenon that is related to the physical characteristics of the surface layer in concentrated contact and that may have relevance to this paper is a creep-type phenomenon observed by the discussor in tests on frictional Hertzian contacts within the region of no gross slip (ref. 47). This phenomenon can be modeled, as shown in the reference, by including a viscous element in Mindlin's model.

#### LECTURER'S CLOSURE

I wish to thank Professor Seireg for calling attention to several additional references. Johnson's paper (ref. 46) discusses the asperities theory of microslip and friction due to Tabor; this paper was apparently unknown even to Nayak and Paul (ref. 40).

While my intention was to write a comprehensive and critical survey of microslip and related contact phenomena, by the time I had covered cylindrical contacts, limitations of space and time forced me to compromise on this ideal intent. In particular, the discussion of the asperities theory of microslip phenomena and friction is confined to the first paragraph of the eighth section and refers only to the paper by Nayak and Paul. One can do nothing but agree with Professor Seireg regarding the inadequacy of such treatment of what promises to be an important advance from the classical Hertz theory.

## REFERENCES

1. CARTER, F. W.: On the Action of a Locomotive Driving Wheel. *Proc. Roy. Soc. (London)*, vol. A112, 1926, pp. 151-157.
2. PORITSKY, H.: Stresses and Deflections of Cylindrical Bodies in Contact. *J. Appl. Mech.*, ASME, vol. 17, 1950, pp. 191-201.
3. CAIN, B. S.: Discussion of reference 2. *J. Appl. Mech.*, ASME, vol. 17, 1950, pp. 465-466.
4. JOHNSON, K. L.: Tangential Traction and Microslip in Rolling Contact. *Rolling Contact Phenomena*, J. B. Bidwell, ed., Elsevier Press, 1962, pp. 6-28.
5. DE PATER, A. D.: On the Reciprocal Pressure Between Two Elastic Bodies. *Rolling Contact Phenomena*, J. B. Bidwell, ed., Elsevier Press, 1962, pp. 29-75.
6. BENTALL, R. H.; AND JOHNSON, K. L.: Slip in Rolling Contact of Two Dissimilar Elastic Rollers. *Int. J. Mech. Sci.*, vol. 9, 1967, pp. 389-404.
7. MINDLIN, R. D.: Compliance of Elastic Bodies in Contact. *J. Appl. Mech.*, ASME, vol. 16, 1949, pp. 259-268.
8. MOW, V. C.; CHOW, P. L.; AND LING, F. F.: Microslip Between Contacting Paraboloids. *J. Appl. Mech.*, ASME, vol. 34E, 1967, pp. 321-328.
9. DESOYER, K.: *Strenge Lösung für die Rollenden Reibung*. München, 1947 (see also ref. 10).
10. BUFLER, H.: Zur Theorie der Rollenden Reibung. *Ing. Arch.*, vol. 27, 1959, p. 137 (Desoyer's results are summarized by Bufler, who also corrects certain sign errors).
11. SÖHNGEN, H.: Zur Theorie der Endlichen Hilbert Transformation. *Math. Zeitschr.*, vol. 66, 1954, pp. 31-51.
12. JOHNSON, K. L.: The Effect of a Tangential Contact Force upon the Rolling Motion of an Elastic Sphere on a Plane. *J. Appl. Mech.*, ASME, vol. 25, 1958, pp. 339-346.
13. CATTANEO, C.: Sul Contatto di due Corpi Elastici. *Accademia de Lincci, Rendiconti*, ser. 6, vol. 27, 1938, pp. 342-348, 434-440, 474-478.
14. JOHNSON, K. L.: The Effect of Spin upon the Rolling Motion of an Elastic Sphere on a Plane. *J. Appl. Mech.*, ASME, vol. 25, 1958, pp. 332-338.
15. JOHNSON, K. L.: The Influence of Elastic Deformation upon the Motion of a Ball Rolling Between Two Surfaces. *Proc. Inst. Mech. Engrs.*, vol. 173, 1959, pp. 795-810.
16. VERMEULEN, P. J.; AND JOHNSON, K. L.: Contact of Non-Spherical Elastic Bodies Transmitting Tangential Forces. *J. Appl. Mech.*, ASME, vol. 31E, 1964, p. 337.
17. HALLING, J.: Microslip between a Rolling Element and Its Track Arising from Geometric Conformity and Applied Surface Traction. *J. Mech. Eng. Sci.*, vol. 6, 1964, pp. 64-73.
18. BROTHERS, B. G.; AND HALLING, J.: Effect of Geometric Conformity on the Slip and Wear in the Contact Region. *Proc. Inst. Mech. Engrs.*, vol. 179, pt. 3J, 1964-65, pp. 134-144.
19. HALLING, J.; AND BROTHERS, B. G.: Wear Due to Microslip Between a Rolling Body and Its Track. *ASME Paper 64-Lub-30*, 1964.
20. HALLING, J.: The Microslip between a Ball and Its Track in Ball-Thrust Bearings. *J. Basic Eng.*, ASME, vol. 88D, 1966, pp. 213-220.
21. DRUTOWSKI, R. C.: The Role of Microslip for Freely Rolling Bodies. *J. Basic Eng.*, ASME, vol. 87D, 1965, pp. 724-734.
22. PORITSKY, H.; HEWLETT, C. W.; AND COLEMAN, R. E.: Sliding Friction of Ball Bearings of the Pivot Type. *J. Appl. Mech.*, ASME, vol. 14, 1947, pp. 261-265.
23. REICHENBACH, G. S.: The Importance of Spinning Friction in Thrust-Carrying Ball Bearings. *J. Basic Eng.*, ASME, vol. 82D, 1960, pp. 295-301.



24. DIETRICH, M. W.; PARKER, R. J.; ZARETSKY, E. V.; AND ANDERSON, W. J.: Contact Conformity Effects on Spinning Torque and Friction. *J. Lub. Tech., ASME*, vol. 91F, 1969.
25. FLOM, D. G.; MULLIN, J. V.; AND SNEDIKER, D. K.: Discussion of reference 24. *J. Lub. Tech., ASME*, vol. 91F, 1969.
26. HAINES, D. J.; AND OLLERTON, E.: Contact Stress Distributions on Elliptical Contact Surfaces Subjected to Radial and Tangential Forces. *Proc. Inst. Mech. Engrs.*, vol. 177, 1963, p. 95.
27. HAINES, D. J.: Contact Stresses in Flat Elliptical Contact Surfaces Which Support Radial and Shearing Forces During Sliding. *Proc. Lub. Conv., Inst. Mech. Engrs.*, 1965.
28. KALKER, J. J.: The Tangential Force Transmitted by Two Elastic Bodies Rolling Over Each Other with Pure Creepage. *Wear*, vol. 11, 1968, pp. 421-430.
29. KALKER, J. J.: A Strip Theory for Rolling with Slip and Spin. *Proc. Konink Nederf. Akad. von Utschappen*, ser. B, no. 70, 1967, pp. 10-62.
30. LOVE, A. E. H.: *Treatise on the Mathematical Theory of Elasticity*. Fourth ed., Cambridge Univ. Press, 1927.
31. MINDLIN, R. D.; AND DERESIEWICZ, H.: Elastic Spheres in Contact Under Varying Oblique Forces. *J. Appl. Mech., ASME*, vol. 20, 1953, pp. 327-344.
32. MINDLIN, R. D.; MASON, W. P.; OSMER, T. F.; AND DERESIEWICZ, H.: Effects of Oscillating Tangential Forces on the Contact Surfaces of Elastic Spheres. *Proc. 1st U.S. Nat. Congr. Appl. Mech.*, 1951, pp. 203-208.
33. KALKER, J. J.: The Transmission of Force and Couple between Two Elastically Similar Rolling Spheres. *Koninkl. Ned. Akad. Uetenschap.*, ser. B, vol. 67, 1964, pp. 135-177.
34. PAPKOVICH: *Comptes Rendue (Paris)*, vol. 195, 1932, pp. 513-515, pp. 754-756.
35. NEUBER, H.: *Kerbspannungslehre*. Second ed., Springer, Berlin, 1958.
36. WHITTAKER, E. T.; AND WATSON, G. N.: *Modern Analysis*. Fourth ed., Cambridge Univ. Press, 1958, ch. 23.
37. KALKER, J. J.: Rolling with Slip and Spin in the Presence of Dry Friction. *Wear*, vol. 9, 1966, pp. 20-38.
38. GALIN, L. A.: Contact Problems in the Theory of Elasticity. Translated from Russian, I. N. Snedon, ed., North Carolina State College, 1961, ch. 2, sec. 8.
39. DORNOROVICH, V. I.: Three-Dimensional Contact Problems, *Theory of Elasticity (in Russian)*, Minsk, 1958.
40. NAYAK, P. R.; AND PAUL, I. L.: A New Theory of Contact. Rept. DSR-76109-7, Eng. Proj. Lab., Dept. Mech. Eng., M.I.T., Cambridge, Mass., April 1968.
41. REITAN, D. K.; AND HIGGINS, T. J.: Calculation of the Electrical Capacitance of a Cube. *J. Appl. Phys.*, vol. 22, 1951, pp. 223-226.
42. MEI, K.; AND VAN BLADEL, J.: Antenna Propagation 11. *IEEE Trans.*, 1963, pp. 52-56.
43. VAN BLADEL, J.: *Electromagnetic Fields*. McGraw-Hill, 1964, pp. 95, 144.
44. JOHNSON, K. L.; AND CAMERON, R.: Shear Behavior of Elastohydrodynamic Oil Films at High Rolling Contact Pressures. *Proc. Inst. Mech. Engrs.*, vol. 182, no. 14, 1967-68, pp. 307-330.
45. TALLIAN, T. E.; CHIU, Y. P.; HUTTENLOCHER, D. F.; KAMENSHINE, J. A.; SIBLEY, L. B.; AND SINDLINGER, N. E.: Lubricant Films in Rolling Contact of Rough Surfaces. *ASLE Trans.*, vol. 7, no. 2, April 1964.
46. JOHNSON, K. L.: Surface Interaction Between Two Elastically-Loaded Bodies Under Tangential Forces. *Proc. Roy. Soc. (London)*, vol. A230, 1955, pp. 531-548.
47. SEIREG, A.; AND WEITER, E. J.: Viscoelastic Behavior of Frictional Hertzian Contacts Under Ramp-Type Loads. *Proc. Wear and Lubrication Conv., Inst. Mech. Engrs.*, 1967.

# The Postulate About the Constancy of Scoring Temperature

**HARMEN BLOK**  
**University of Technology**  
**Delft, Holland**

The author has comprehensively reanalyzed numerous sets of published data on the scoring performance of nonadditive mineral oils in gears and simulating testers. This was performed in order to shed more light on the postulate about the constancy of scoring temperature. According to their originators' interpretations, several of these sets of data seemed to invalidate the postulate; but many of these interpretations called for certain refinements, most of which the author has long had available from previous work and now discloses in the form of case studies.

It would appear that most design calculations on the postulate's applicability range are sufficiently wide for counterformal rubbing surfaces for their scoring risk. Also, that the scatter inherent in scoring temperatures, obtainable through the refined methods now available, is not greater than that normally accepted for conventional design data on both the bulk and surface fatigue strength of gear steels. There is still a need, however, for more certainty about both the lower and the upper bound to the applicable scoring temperature range.

One group of refinements discussed herein bears upon flash temperature theory. The corresponding generalizations of this theory are given.

Another group of refinements relates to the prediction of bulk temperature. In general, this is as important as the flash temperature in estimating the maximum temperature in the conjunction zone of two rubbing surfaces for which the scoring risk is to be assessed. Prompted by the lack of a sufficiently refined method for such a prediction, the author outlines a thermal-network theory that for gear transmissions had long proved its worth in his own work.

Despite a prolonged search by the author and others, no single scoring mechanism or combination of such mechanisms is known that might satisfactorily explain the wide applicability of the present postulate. Therefore, various conceivable mechanisms are submitted for discussion.

**T**HE ESTABLISHMENT OF THE FLASH TEMPERATURE THEORY (refs. 1 to 3) opened up an avenue for estimating the temperatures generated by the frictional heat in the conjunction zone between counterformal rubbing

surfaces. In fact, denoting the maximum conjunction temperature by  $T_c$ , one may put:

$$T_c = T_b + T_f \quad (1a)$$

where  $T_b$  represents the bulk temperature. This is the temperature representative of the fairly uniform level of those parts of the temperature fields in the rubbing bodies that do not lie too close to the conjunction zone.  $T_f$  stands for the maximum flash temperature occurring in that zone and may be calculated from the following basic equation:

$$T_f = 1.11 \frac{fW |V_1 - V_2|}{b_1 \sqrt{V_1} + b_2 \sqrt{V_2}} \cdot \frac{1}{\sqrt{w}} \quad (1b)$$

In the particular, though not infrequent, case where the thermal contact coefficients,  $b_1$  and  $b_2$ , of the rubbing materials 1 and 2 are the same, they may be replaced by their common value,  $b$ . Hence, equation (1b) reduced to<sup>1</sup>:

$$T_f = 1.11fW |\sqrt{V_1} - \sqrt{V_2}| \cdot (b\sqrt{w})^{-1} \quad (2a)$$

or to one of the following variants:

$$T_f = 0.62fW^{3/4} |\sqrt{V_1} - \sqrt{V_2}| \cdot R^{-1/4} \cdot E_r^{1/4} \cdot b^{-1} \quad (2b)$$

and

$$T_f = 2.45f\sigma_H^{3/2} |\sqrt{V_1} - \sqrt{V_2}| \cdot R^{1/2} \cdot E_r^{-1/2} \cdot b^{-1} \quad (2c)$$

where the numerical factors, 1.11, 0.62, and 2.45, respectively, are valid for a semielliptic or, Hertzian distribution of the frictional heat over the instantaneous width,  $w$ , of the supposedly band-shaped conjunction zone. Apart from  $T_f$  and  $w$ , the following symbols have been introduced:

- $f$ , the instantaneous coefficient of friction;
- $W$ , the instantaneous actual load per unit width of the face or track, including the dynamic load increment, whether this be positive or negative;
- $V_1$  and  $V_2$ , the instantaneous tangential velocities (perpendicular to the line of action in the case of meshing tooth faces) of the two rubbing surfaces relative to the conjunction zone;
- $\sigma_H$ , the instantaneous maximum pressure according to Hertz's theory of elastic contact, at the given load,  $W$ ;
- $R$ , the instantaneous radius of conformity, as defined by:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \quad (3a)$$

<sup>1</sup> Equation (1b), like equations (2a), (2b), and (2c), is valid for cases where the bulk temperatures,  $T_{b,1}$  and  $T_{b,2}$ , of the two rubbing bodies are equal. For a generalized equation for the flash temperature component in cases where the two bulk temperatures are unequal, refer to Appendix C (see also eq. (4b) of the present text).



where  $R_1$  and  $R_2$  denote the radii of curvature of the two rubbing surfaces in their unloaded and undeformed state;  
 $E_r$ , the reduced modulus of elasticity, as defined by

$$\frac{1}{E_r} = \frac{1}{2} \left( \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \right) \quad (3b)$$

where  $E_1$  and  $E_2$  stand for the moduli of elasticity and  $\nu_1$  and  $\nu_2$  for the Poisson's ratios of the rubbing materials 1 and 2, respectively;

$b_1$  and  $b_2$ , the thermal contact coefficients of the rubbing materials 1 and 2, respectively.  $b$ , the only variable not used in conventional gear design, is defined by  $b = (k\rho c)^{1/2}$ , where  $k$  stands for heat conductivity,  $\rho$  for density, and  $c$  for specific heat per unit mass, so that  $\rho c$  denotes the specific heat per unit volume.<sup>2</sup>

The inherent limitations were in the empirical criteria of Hofer (refs. 4 and 5) and the better known *PV*-criterion of Almen (refs. 6 to 8) for design calculations for the scoring of gear teeth. As a result, the author applied equations (1a) and (1b) for estimating the maximum conjunction temperatures reached at incipient scoring in several gear testing machines, as well as in actual gear practice. The author thus arrived at the conclusion that, at least with nonadditive mineral oils, each combination of oil and rubbing materials considered could be characterized by a maximum conjunction temperature. To the accuracy of the estimates and within the fairly wide range of operating conditions already explored at that time, each such scoring temperature appeared to be constant. Since the author recognized the constancy of such characteristic scoring temperatures as a significant design tool, quite apart from their intrinsic physical and/or chemical explanation, he thereupon enunciated the postulate about this constancy (ref. 9).

As is evidenced, among other things, by modern textbooks on gear design (refs. 10 and 11) and also by the extensive exploratory work of Lemanski (ref. 12), the postulate, in conjunction with the flash temperature theory, has been applied increasingly by gear designers. Furthermore, promising attempts have been made to apply the postulate to other kinds of counterformal rubbing surfaces such as cams and tappets (refs. 13 and 14), including the selection of their lubricants, in order that they may be expected to operate free from scoring.

Of course, the author has realized that limitations of the applicability

<sup>2</sup> For the steels normally used for gears, the thermal contact coefficient,  $b$  ranges from about 12 to 16 kgf cm<sup>-1</sup> sec<sup>-1/2</sup> C<sup>-1</sup>, or about  $1.2 \times 10^4$  to  $1.6 \times 10^4$  Jm<sup>-1</sup> sec<sup>-1/2</sup> C<sup>-1</sup>, or about 67 to 90 lbf in<sup>-1</sup> sec<sup>-1/2</sup> C<sup>-1</sup> = 37 to 50 lbf in<sup>-1</sup> sec<sup>-1/2</sup> F<sup>-1</sup>. Note that equations (1b), (2a), (2b), and (2c) are valid in any consistent system of units, but only in such a system.

of the postulate might show up which are narrower than those specified originally in 1939. It might, for instance, prove that as operating conditions are hydrodynamically more favorable, the approximation inherent in the constancy of scoring temperature becomes cruder and cruder. But in this particular respect, the experimental evidence available (refs. 15 and 16) is not yet thought to be sufficiently enlightening.

The ultimate verification, that is, using gears as the test specimens, has always proved most difficult. In fact, the data normally obtained on the various standard gear machines in regular use for scoring tests (i.e., those in terms of the scoring load of a given mineral oil as a function of the rotational speed and/or of the temperature of the oil as supplied to the gears), do not suffice for verifying the constancy or nonconstancy of scoring temperature in a really decisive manner. For this purpose additional knowledge, which is not normally available for such machines, is required, such as:

(1) The bulk temperature,  $T_b$  (see eq. (1a)), of the test gears at incipient scoring. In general, this temperature cannot be equal to the known temperature of the oil as supplied to the gears;

(2) The dynamic tooth load increments. While these increments will undoubtedly be positive during one or more portions of the meshing cycle, they must be negative during the remaining portions.<sup>3</sup> The possibility, not even to be excluded, is that the increment concerned happens to be negative at the particular meshing position where incipient scoring occurs;

(3) The variation of the coefficient of tooth friction from one meshing position to another, i.e., as dependent on certain viscosity characteristics of the oil and on the variations in the tangential velocities along the tooth faces and in the actual tooth load, including the dynamic increments.

Although notable advances have recently been made in the above-mentioned pieces of additional knowledge on gears, the verification of the postulate still depends for the most part on experiments in simulating testers, such as disk machines, where some of the additional knowledge required is more readily obtained. Of course, it may be questioned whether the results of such "simulating verifications" may be applied to actual gear practice.

On the other hand, for the design of gears there is no real need for utmost precision in the verification of the constancy of scoring temperature. This is so because of two major reasons:

---

<sup>3</sup> The average of the dynamic tooth load increments over every meshing cycle must be zero. After all, the dynamic load increments are by definition to be superimposed upon the nominal tooth load which follows from the given power transmitted and the given rotational speed, and which thus represents the exact average of the actual tooth loads over each meshing cycle.



1. Just as in experiments on tooth breakage and pitting, the scatter in scoring loads on gears is far from negligible, even on the high-precision gears that are used as test specimens in the various standard machines. It would be unreasonable to demand a stricter constancy and less scatter of scoring temperatures than is normally considered acceptable for basic design data on tooth breakage and pitting.
2. The ultimate aim of knowledge about scoring temperatures, just as that about tooth breakage and pitting, lies in the design of gears. In the first place, gears to be designed for actual practice will rarely have to be made as accurate as the test gears considered. Therefore, as with the scatter in data on tooth breakage and pitting, the designer will not be handicapped as long as the scatter in scoring data is not greater. Furthermore, in the design stage, because of the uncertainties inherent in the manufacturing tolerances, even exact basic data cannot possibly result in really accurate predictions about the performance of the gears yet to be manufactured and assembled.

It is the main purpose of the present paper to shed some more light upon the various questions about the applicability of the present scoring temperature postulate.

The next section illustrates sources of, and methods of eliminating, the various uncertainties and inaccuracies in experimental verifications of the applicability of the present postulate.

"Refinements in the Theoretical Prediction of Conjunction Temperatures" treats various refinements in the theoretical prediction of both components of conjunction temperature (see eq. (1a)), the flash temperature component, and the bulk temperature component. Certain of the generalizations needed in some of the refinements are dealt with in appendixes.

"Scoring Mechanisms" describes various mechanisms that are conceivable as the basic causes of scoring. Since this subject still remains in a rather nebulous stage, it is presented mainly for discussion.

The final section summarizes the major conclusions and recommendations.

#### EXPERIMENTAL VERIFICATIONS OF THE PRESENT POSTULATE

##### General Considerations

Verification of the present postulate is tantamount to verification of the constancy of the maximum conjunction temperature at incipient scoring in as wide a range of operating conditions as possible. Various methods for experimental verification are conceivable, and in all of them one may benefit from the distinction, as expressed in equation (1a), between the two components of the maximum conjunction temperature,  $T_c$ , i.e., the bulk temperature component,  $T_b$ , and the maximum,  $T_f$ , of the flash



temperature. These methods will be illustrated by means of case studies made of a sequence of sets of experiments by various investigators.

The first method is that of direct measurement where, practically speaking, one is committed to the use of the principle of the so-called "dynamic thermocouple." This principle is characterized by the feature that the two rubbing bodies are themselves used as the elements of the thermocouple required.

Limitations and difficulties are inherent in the method of the dynamic thermocouple in that, to achieve a suitable thermoelectric potential, one is not free in the choice of the two rubbing materials. For instance, it proves difficult to find thermoelectrically suitable combinations among the various kinds of steels feasible for gear teeth, particularly for highly loaded ones.<sup>4</sup> This difficulty may explain why it is doubtful whether the results thus obtained on the conjunction temperatures at incipient scoring are sufficiently representative of the different combinations of steels that are commonly found in actual gear practice. Such scoring results have been published by Meng (ref. 17) on an ordinary disk machine having continuously rotating disks and by Fleck (ref. 18) on an oscillating—disk machine.<sup>5</sup>

Since the flash temperature component and, therefore, the conjunction temperature, varies over the width of the conjunction zone, i.e., in the direction of the motion, the thermoelectric potential measured is representative of some kind of average of these temperatures over the width concerned. Judging from the theories of Emmons (ref. 19) and Shu (ref. 20) about such "extended thermocouples," in both Meng's and Fleck's experiments this average must have been very close to the arithmetic average.

Investigations by various experimenters on a variety of testers will now be discussed. This will be done in the form of case studies in which, among other things, the need for certain refinements in the flash temperature theory and its application will be set forth.

#### Case Study of the Disk Experiments of Meng

Meng (ref. 17) found a nearly constant scoring temperature for the one lubricant, a nonadditive mineral oil, tested by him. But this temperature was given in terms of the average conjunction temperature. Knowing, however, the bulk temperature measured by Meng in his various tests,

---

<sup>4</sup> For other difficulties, see the author's discussion on a paper by Niemann and Lechner (ref. 16) and their closure.

<sup>5</sup> Among Niemann and Lechner's (ref. 16) extensive thermoelectric test series on a gear machine, there were only two tests where load was increased up to incipient scoring, so that the two ensuing scoring temperatures could be measured. However, since the test oil contained an EP additive, these two tests do not lend themselves to verifying the postulate. Moreover, the combination of steels used in one of these tests was far from representative for actual gear practice.

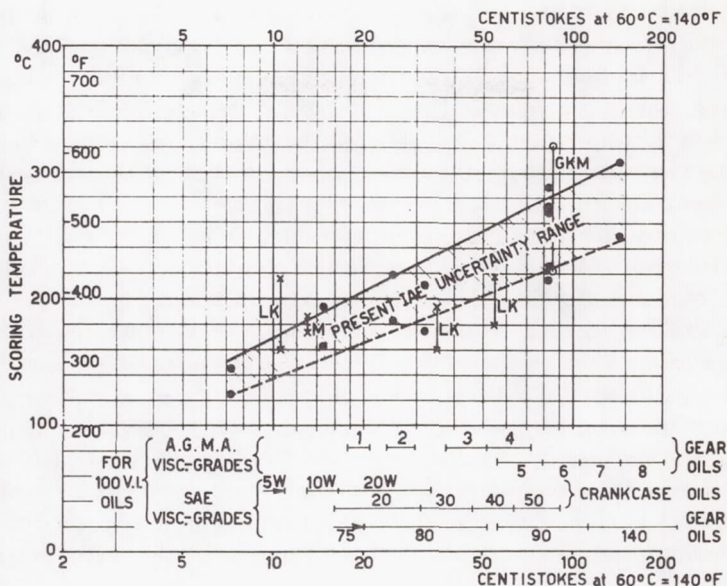


FIGURE 1.—Survey chart of scoring temperatures for nonadditive mineral oils of various viscosity grades.

it was an easy matter to convert each of his scoring temperatures to a corresponding maximum conjunction temperature. In fact, for a semielliptic distribution of the frictional heat, the distribution of the flash temperature component is such that its maximum is 1.45 times its average.

It was thus found that for Meng's mineral oil, his test results corroborated the author's postulate in that the maximum temperatures at incipient scoring were satisfactorily constant. The range of these converted temperatures has been indicated in figure 1 and falls within the uncertainty range for the various nonadditive mineral oils tested previously in the IAE  $3\frac{1}{4}$  in.-centres Standard Spur Gear Rig.<sup>6</sup>

#### Case Study of the Disk Experiments of Fleck

Fleck (ref. 18) verified both the flash temperature formula and the postulate, also by means of the dynamic thermocouple method. Because of the oscillating motion in his disk machine, the conjunction

<sup>6</sup> Figure 1 is an amended form of the survey diagram originally given by the author (ref. 21). The amendment was given later when better estimates had become available for the bulk temperatures of the gears in the IAE Rig for the various load increments. (See the figure included in the discussion by Blok on a paper by Leach and Kelley ref. 22.)



temperatures varied cyclically, as was indicated on his oscillographic records. By deducting the constant surface temperature of the disks, which was measured at or near the inlet to the conjunction zone, he assessed the flash temperature component which, of course, varied cyclically also. Throughout each cycle he found this component to be about 30 percent higher than the one calculated, with the measured coefficients of friction that equally varied cyclically, from the flash temperature equation (2a). Before one can be sure of the reality of this difference and thus of a violation of the flash temperature theory, however, at least the following two points should be considered carefully.

First, Fleck assumed the bulk temperature of the disks equal to the aforementioned inlet temperature at their surface. However, since there was an abundant jet supply of the lubricating oil on either side of the conjunction zone, it would appear that the bulk temperature may well have been sensibly higher than assumed by Fleck. According to a first estimate of the order of magnitude of the difference concerned (see Appendix A), the ensuing correction might well bring Fleck's flash temperature component much closer to the one following from equation (2a).

While the above comment is more or less specific of Fleck's procedure, the following one also applies both to many other disk machines and to gears, either to standard testers or, even more so, to actual gear practice.

Even with the most careful alignment, edge effects are unavoidable as is evidenced by the fact that scoring does not usually occur all over the width of the disks or tooth faces, but tends to start at an edge. Thus the conjunction temperatures may then be assumed not to be uniform and to be highest at the edge concerned. This is another reason why the conjunction temperature as measured by the dynamic thermocouple technique, and thus the flash temperature derived therefrom by Fleck, will be on the high side as compared with that calculated from equation (2a).

This reason makes it even more doubtful whether Fleck's interpretation of his results has provided definite proof for his claim that, in his range of operating conditions and with all of the four nonadditive mineral oils, the postulate was indeed violated.

The above comments, like others to follow, have served to illustrate that one has to devote considerable attention to the selection of suitable measuring methods to verify the flash temperature theory, and even more so when the postulate is at stake. Furthermore, the comparatively great sensitivity of the interpretation of the measurements obtained towards certain details, which at first may look so minute as to be negligible, calls for the utmost caution. For instance, as is shown in Appendix B and illustrated in figure 2, a minute misalignment of disks or tooth faces of the usual widths may easily alter the hypothetically parallel band occupied by the conjunction area into a tapered one having a maximum width,



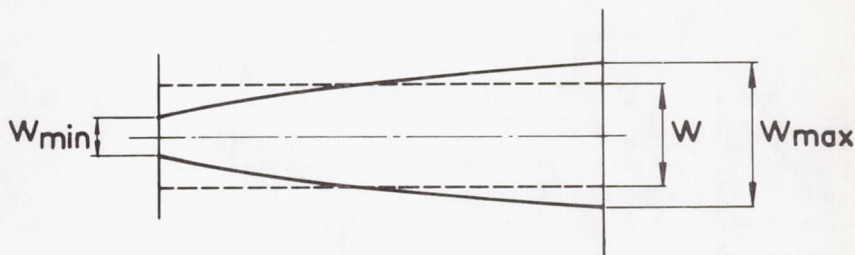


FIGURE 2.—Tapering of the band occupied by the conjunction area as caused by misalignment.

$w_{max}$ , that is as much as 10 percent greater than the normally uniform width,  $w$ , at the given load (for the tapering, see fig. 2).

It is also shown in Appendix B that the actual maximum conjunction temperature is proportional to  $(w_{max}/w)^{3/2}$ . Accordingly, even with the minute misalignment specified above by  $w_{max}/w = 1.10$ , the actual maximum conjunction temperature is as much as 15.4 percent greater than the nominal one that is calculated for uniform width at the same load from equation (2a).

The conclusion is that for the present verification purposes one should eliminate the misalignment difficulty altogether, instead of correcting the results through posteriori estimates like those outlined above. The best technique of elimination known to the author is the one consisting of the use of crowned disks, as adopted by Kelley (refs. 23 and 24).

At least for the time being, the author puts more faith in Kelley's interpretation of his tests in which the bulk temperature component was measured while the flash temperature component was calculated from measurements of the various influential factors in the flash temperature equation (2a). The results thus obtained corroborated the postulate in quite a wide range of operating conditions. Finally, for Kelley's scoring temperatures for three nonadditive mineral oils, and their reasonable amount of scatter, the reader is referred to their results in the survey diagram of figure 1. (The three vertical bars marked by Leach and Kelley (ref. 22) indicate the three scatter ranges.)

#### Case Study of Gear Experiments of Blok and of Niemann and Lechner

Let us now turn to attempts at verifying the flash temperature theory and/or the postulate in actual gears. Perhaps the first such attempt was made by the author (ref. 2) utilizing the dynamic thermocouple method. The epicyclic gear machine especially set up for the purpose had one broad planet meshing with two narrow sun gears that were stationary while two different and thermoelectrically suitable steels were used. So, fixed leads could be used for measuring the thermoelectric potential generated between the two sets of meshing teeth.

Since at the time no suitable oscillographs were available, the author had to confine himself to the average of the thermoelectric potential over each meshing cycle, as measured on a plain millivoltmeter having a slow response. Therefore, the success achieved was not complete in that the flash temperature theory could be verified only in the overall way implied by the averaging measurements. Moreover, the operating conditions could not be made sufficiently severe to induce scuffing so that the postulate could not be checked at all.

Later, Niemann and Lechner (ref. 16), using also the method of the dynamic thermocouple, did succeed in obtaining oscillographic records of the variation of the conjunction temperature during meshing cycles. As in the author's above-mentioned experiments on gears and those of Fleck (ref. 18) on oscillating disks, the fact that thermoelectric potentials could indeed be measured is indicative of the fact that fluid-film lubrication may at its best have been only partial in the operating range covered.

In a qualitative way the oscillograms of the conjunction temperature came up to expectations on the basis of the flash temperature theory as applied to the successive meshing positions during each meshing cycle. However, the data provided in Niemann and Lechner's paper did not suffice to put the author in a position to verify the flash temperature theory also in a quantitative way. This is an unfortunate circumstance since in their own verification Niemann and Lechner (see their fig. 16) compared the conjunction temperatures as measured with the ones calculated by them from the Blok-Kelley formula. Thus they calculated the flash temperature component from this formula on the tacit assumption that the tooth load would have been distributed uniformly over the width of the teeth. On the other hand, they pointed out that, as in the tests of Borsoff (ref. 15), the scoring load per unit width decreased with increasing tooth width. This feature is indicative of some amount of maldistribution, and thus of concentration, of the tooth load. Further, in keeping with the derivation of the flash temperature equation (2a), where  $w$  stands for the actual load, Niemann and Lechner should rather have accounted for the dynamic load increments as well as the load concentration (see Appendix B), instead of substituting the nominal load.

Because of the above arguments, the author does not share Niemann and Lechner's opinion that the flash temperature theory did not hold good in their tests.<sup>7</sup>

In their reply to the discussion on their paper, Niemann and Lechner gave in their figure 18 the conjunction temperatures at incipient scoring as measured by them for various rotational speeds of their test gears. Since these tests were performed with an EP oil, these temperatures

---

<sup>7</sup> For additional points raised by the author and for Niemann and Lechner's reply, see pp. 641-654 following their paper (ref. 16).

cannot serve for testing the postulate about the constancy of scoring temperatures, this postulate being confined to nonadditive mineral oils.

#### Case Study on Gear Experiments of Lechner

Later, Lechner (ref. 25) published results that did relate to a non-additive mineral oil. (See fig. 3, which has been redrawn from Lechner's fig. 1.)

In contrast to the above-discussed tests of Niemann and Lechner (ref. 16), Lechner did not measure the conjunction temperatures,  $T_{c,s}$ , at incipient scoring, but only the bulk temperatures at the center of the teeth, which may be taken identical with the bulk temperatures,  $T_b$ . In fact, he calculated the flash temperature component,  $T_f$ , from the relevant equation (2a) and superimposed it on the bulk temperature,  $T_b$  (see eq. (1a)). His conclusions were as follows:

1. The postulate about the constancy of conjunction temperature at incipient scoring did not hold good in his tests. (see the upward trend of the  $T_{c,s}$  curve in fig. 3.)
2. At least up to a circumferential velocity of about 25 m/sec of Lechner's test gears, the bulk temperature was reasonably constant (see the lower curve in fig. 3) and might be taken as a criterion for the occurrence of incipient scoring.

Regarding the first conclusion, however, the following comments are worthy of note.

In the first place, in his calculations of the flash temperature component Lechner assumed the value of the coefficient,  $f$ , of tooth friction to be constant, i. e., independent of both the meshing position and the tan-

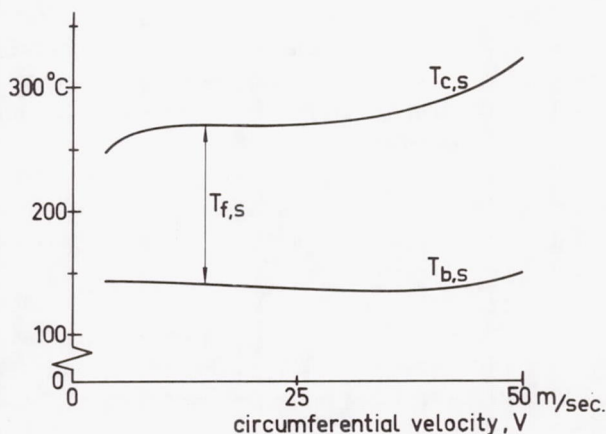


FIGURE 3.—Scoring and central tooth temperatures determined by Lechner (ref. 25) for a nonadditive mineral oil on his gear testing machine.



gential velocities of the meshing tooth faces or, say, of the circumferential velocity, which in his tests (fig. 3) ranged from about 4 to 50 m/sec. However, at the lower end of this velocity range it may be expected that, even at loads lower than the scoring loads concerned, lubrication will have been only of the partially hydrodynamic type. In fact, in their previous tests on presumably the same gear testing machine, Niemann and Lechner would otherwise not have been able to measure conjunction temperatures up to circumferential velocities as high as 30 m/sec and at loads lower than those at incipient scoring. So, in reality the coefficient of tooth friction may safely be assumed to have decreased with increasing circumferential velocity, again up to at least something like 30 m/sec.<sup>8</sup>

Now, the ensuing correction of the flash temperature component  $T_f$ , and thus for the conjunction temperature,  $T_{c,s}$  at incipient scoring, would tend to make the actual trend of the latter temperature more constant than that in figure 3, at least up to about 30 m/sec.

Summarizing, the author so far fails to see why, in the very same velocity range, his postulate could not be upheld and should be replaced by Lechner's postulate as expressed by his above second conclusion.

Furthermore, it might well prove difficult, if not impossible, to generalize Lechner's postulate so as to make it applicable to actual gear practice. In fact, a comparatively small pinion tends in an overall sense to become appreciably hotter than the meshing gear, or, say, the former's bulk temperature tends to be appreciably higher. In the case of Lechner's postulate, this phenomenon leaves the designer with the rather awkward choice of a bulk temperature representative of the meshing zone or, say, of scoring. In the case of the author's postulate, however, there would be no such choice, the conjunction temperature being common to the pinion and the meshing gear.

It might prove even harder to explain why some critical bulk temperature, such as postulated by Lechner, could be decisive in determining the particular meshing position in the meshing zone where scoring starts on a given pair of meshing teeth.

Let us now discuss the remaining portion of Lechner's range of circumferential velocities, that from about 30 m/sec upward, where the trend of his conjunction temperature,  $T_{c,s}$ , at incipient scoring is rather steep (compare fig. 3).

Because of its steepness it might not prove possible to reduce the relevant portion of Lechner's  $T_{c,s}$  curve to a more or less constant level solely by correcting for the variation of the coefficient of tooth friction with circumferential velocity or, more precisely, with the tangential velocities

---

<sup>8</sup> For estimating this variation of the coefficient of tooth friction with circumferential velocity or, more precisely, with the tangential velocities of the meshing tooth faces in the successive meshing positions, see for instance equation (11.5) of Kelley and Leman-ski (ref. 24).

of the tooth faces at the critical meshing position where incipient scoring will have occurred. It is even conceivable that, in the present range of comparatively high velocities, lubrication might have approached so closely to the full fluid-film type; as a result, with increasing velocity, the coefficient of tooth friction would go up. If this phenomenon should have occurred, it would, by itself, tend to make the corrected trend of the conjunction temperatures of incipient scoring even steeper than the original one of Lechner. But it might well prove that additional corrections of other kinds could more than offset the upward trend of Lechner's curve, or that of the curve obtained after having corrected only for a possibly increasing coefficient of tooth friction.

It would appear that several such additional corrections may still have to be applied simultaneously to Lechner's curve. One such correction is that for maldistribution of the tooth load (see fig. 2 and Appendix B). This correction will vary with load and thus with the circumferential velocity, the load concerned being identical to the scoring load.

Another correction is the one relating to the dynamic tooth load increments which will also vary with circumferential velocity. In this connection it is pointed out that the critical meshing position at which incipient scoring will have occurred in Lechner's experiments may not invariably have been the same throughout his range of circumferential velocities. If so, a third correction would also be necessary. In fact, at some transitional circumferential velocity within that range, a change-over from one critical meshing position to another might well alter the trend of the corrected  $T_{c,s}$  curve quite appreciably as compared with Lechner's uncorrected one.

Further, it may well be that in Lechner's experiments the bulk temperature of the pinion differed too much from that of the gear wheel<sup>9</sup> to permit the use of the original flash temperature equation (2a), which after all was developed under the assumption of a bulk temperature common to pinion and gear. For the correction then required, in the form of a generalization of the flash temperature equation for two unequal bulk temperatures, the reader is referred to Appendix C.

It remains to be seen whether or not the fully corrected trend of the  $T_{c,s}$ - $v$  relationship in Lechner's experiments would lend better support to the author's postulate.<sup>10</sup> All in all, for the time being, Lechner's evi-

<sup>9</sup> The gear ratio in Lechner's gear tester amounted to 1.5 so that in an overall sense the pinion, being distinctly smaller than the meshing gear, might have been noticeably hotter than the gear.

<sup>10</sup> Corrections of still other kinds are conceivable, e.g., that of Ling and Mow (ref. 26) or Korovchinski (ref. 27) for thermal distortion of the two rubbing surfaces in their conjunction zone, and that of Ling and Rice (ref. 28) for the temperature dependence of heat conductivity,  $k$ , and specific heat,  $\rho c$ , per unit volume, and thus of the thermal contact coefficient,  $b = (k\rho c)^{1/2}$ . It is thought, however, that these two corrections are not really important in the range of Lechner's tests.



dence in the form of his  $T_{c,s}-v$  curve in figure 3 is not considered sufficiently convincing for discrediting the author's postulate.

#### Case Study on Gear Experiments of Terauchi and Miayo

Even prior to Niemann and Lechner, Terauchi and Miayo (ref. 29) obtained oscillograms of the conjunction temperatures by means of the dynamic thermocouple method, which was applied to a constantan pinion and a steel gear. Unfortunately, they did not give any bulk temperatures so that the flash temperature theory could not be verified in a quantitative way. But in a qualitative way their work was still useful in that it illustrated the cyclic variation of the conjunction temperatures to show the pattern to be expected from the flash temperature theory, accounting even for the dynamic load increments and the sharing of the total tooth load that takes place throughout the portion of double tooth-contact during each meshing cycle. Furthermore, they were able to show that there is an optimum in profile modification of the teeth as regard the maximum conjunction temperatures reached. Finally it is noteworthy that the postulate cannot be verified by Terauchi and Miayo's experiments since they did not explore any scoring conditions.

### REFINEMENTS IN THE THEORETICAL PREDICTION OF CONJUNCTION TEMPERATURES

#### General Considerations

As long as it is accepted that conjunction temperatures are significant in the prediction of scoring, it is important that, through an adequate method of design calculation, the designer be put in a position to predict these temperatures sufficiently accurately. This is true even of ranges of operating conditions, if any, where the critical conjunction temperature at incipient scoring would vary with one or more of these conditions, so that the postulate would not hold good there.

The author will confine himself to predictions of the conjunction temperatures on the basis of flash temperature theory. Since any conjunction temperature,  $T_c$ , consists of a bulk temperature component,  $T_b$ , and a flash temperature component,  $T_f$ , either component will have to be predicted individually.

Since the main lines of the prediction of the flash temperature component have already been dealt with in previous papers and also in the foregoing section, i.e., in the form of the basic flash temperature equation and its variants, the author will confine himself in the following subsection to certain refinements in the prediction concerned.

While a fairly well rounded-off theory, including the following refinements, is now available for predicting the flash temperature component, a sufficiently general theory for predicting the bulk temperature component has, to the best of the author's knowledge, not yet been published. In the bulk temperature refinement section, he will give an outline of



such a theory which, for reasons set forth there, will be called "the thermal network theory."

#### Refinements in the Theoretical Prediction of the Flash Temperature Component

Two kinds of refinements will be considered here. The first kind relates to refinements in and generalizations of the flash temperature theory itself and the second kind to refinements in the practical application of this theory for the prediction proper. To begin with, however, the author will discuss one of the most basic assumptions underlying the flash temperature theory—the absence of interfacial temperature jumps in any area of real contact.

*Discussion on the absence of interfacial temperature jumps.*—As long as the basic assumptions underlying the original flash temperature theory (ref. 1) are taken for granted, this theory may be considered as essentially sound.<sup>11</sup> Therefore, we shall first go into the questions of to what extent and in what range one or more of these assumptions break down to such an extent that refinements indeed have to be introduced.

Perhaps the most basic assumption is the absence of temperature jumps across the interface of the two rubbing surfaces. This is tantamount to saying that throughout the conjunction zone, but not generally outside it, the distribution of the temperatures at one of the surfaces is assumed to be identical with that at the mating surface.

The main problem in any theory of flash temperatures, whether the original or a refined one, consists in determining the partitioning, among the two rubbing bodies in their conjunction zone, of the supposedly known heat that is developed in total in that same zone by friction. In the range covered by the original flash temperature theory, the assumption of the absence of interfacial temperature jumps renders the solution of this main problem particularly simple. This is the range specified by the condition that both the Péclet numbers concerned should be sufficiently high.<sup>12</sup> But for gears there is no real need to refine the original flash temperature theory for the range of comparatively small Péclet numbers. The reason being that, on meshing tooth faces, both numbers will be sufficiently high whenever there is even the slightest risk of scoring.<sup>13</sup>

---

<sup>11</sup> For the simple physical model that is representative of the original flash temperature concept, see also another paper of the author (ref. 30).

<sup>12</sup> The two Péclet numbers are  $V_1 w/a_1$  and  $V_2 w/a_2$  where  $V$  denotes the tangential speed of the rubbing surface considered, with respect to the band-shaped conjunction zone,  $w$  the width of this zone, and  $a$  the thermal diffusivity of the rubbing material concerned, where by definition,  $a = k/\rho c$ . Further the subscripts 1 and 2 refer to the rubbing surfaces 1 and 2, respectively.

<sup>13</sup> Several cases where both Péclet numbers are small, or only one of them is small or even zero (the latter case is where the rubbing surface concerned, like the stationary test specimen in the Timken Machine, is at a standstill with respect to the conjunction zone) or where one of them is negative, had already been evaluated by the author by hand calculation when he published the original, elementary theory of flash tempera-

On physical grounds the author is still convinced that the assumption about the absence of interfacial temperature jumps across the conjunction zone is basically correct at any point where contact prevails, even if only instantaneously. In cases, however, where a full fluid film completely separates the two rubbing surfaces, this assumption must be modified. In fact, neither rubbing surface should show a temperature jump with respect to the lubricant particles contacting it directly. Thus in such cases there will in general be a difference in the boundary temperatures of the lubricant across the film, that is, in the temperatures of any pair of points that face each other on the two rubbing surfaces.<sup>14</sup>

Later we shall revert to the above cases where an intervening lubricant film complicates matters quite appreciably in that the presence of a temperature field covering that entire film has to be accounted for. As a start, let us confine ourselves to the cases where the amount of lubricant present in the conjunction zone is so minute that, practically speaking, boundary lubrication will prevail. Now, boundary lubricant films are so thin that any temperature difference across them must be vanishingly small. Accordingly the assumption about the absence of temperature jumps between the two rubbing surfaces in their conjunction zone may here be upheld.

As far as the author is aware, the only experiments where the present assumption has been put to test for rubbing surfaces are those of Ling and Simkins (ref. 35). Since no lubricant was used, contact must have occurred; but, in view of the conformity of the rubbing surfaces used and the comparatively low loads applied, it must have been of a rather disperse nature. That is, in contrast to the much more concentrated contact that will usually occur on the counterformal rubbing surfaces with which we are concerned. The areas of actual contact, which can be referred to as the thermally contacting asperities, must have been comparatively small and have occupied only a very small fraction of the conjunction zone, which constitutes the available area of contact.

Ling and Simkins succeeded in measuring the temperature field in either rubbing body at various points. But the distance between the

---

tures (ref. 1). In 1966 the author's assistant, Mr. De Winter, in a thesis elaborated upon these cases and evaluated them more accurately on a computer. Meanwhile, other investigators have published their numerical results on similar cases (Allen, ref. 31; Cameron et al., ref. 32; and Symm, ref. 33). By comparison with the author's or, still better, with De Winter's more accurate results, it proved that—of Allen's results—only those for a particular case where one of the Péclet numbers is zero deviated strongly from ours and thus are considered wrong. Further our conclusion is that, at least for certain of their cases, there must be errors in the approach followed by Cameron et al. Finally Symm's results corroborate ours.

<sup>14</sup> An appropriate thermal boundary condition, which is based on the basic ideas underlying the flash temperature theory, was suggested by the author in his discussion on a paper by Osterle et al. (ref. 34). Notwithstanding their objection to this boundary condition, it may be proved to be essentially correct.



interface in the conjunction zone and the measuring points closest to this interface was slightly over 0.12 in. (3 mm). So, for either rubbing body they could determine the interface temperature only by extrapolating the temperature field in the bulk material concerned. The two extrapolated interface temperatures thus found proved to be different. But, as Ling and Simkins pointed out, the aforementioned distance was so great relative to the size of the thermally contacting asperities that the temperature difference found may not be taken as indicative of a real temperature jump, that is, a jump on the asperities concerned.

A similar situation had long been recognized by many investigators for the simpler cases of conformal surfaces that do not rub on each other so that both surfaces are stationary with respect to the available contact area, i.e., their conjunction zone (see the references given in ref. 36). It is considered reassuring that, at least in the latter cases, no convincing evidence has ever been presented for the occurrence of real interfacial temperature jumps. Therefore, the author is inclined to maintain his basic assumption about the absence of such a jump.

*Refinements for the temperature dependence of thermal characteristics.*—The first refinement to be considered is the one that may have to be applied when the influence of the temperature dependence of the thermal properties of at least one of the two rubbing materials may no longer be neglected.

The only, albeit combined, thermal characteristic appearing in the original flash temperature equation (2a), is the thermal contact coefficient,  $b$ , which has been assumed constant. It is defined as  $b = (k\rho c)^{1/2}$ ; and, in general, two values of this characteristic need be known,  $b_1$  and  $b_2$ , one for either rubbing material.

As has been shown by Ling and Rice (ref. 28), whenever the temperature variation of  $b_1$  and/or  $b_2$  is sufficiently great in the temperature range to be considered, heat conductivities,  $k_1$  and  $k_2$ , and specific heats,  $(\rho c)_1$  and  $(\rho c)_2$ , no longer appear solely in their combinations,  $b_1$  and  $b_2$ . The more general solution then required for the flash temperature problem proved to be much more complicated and could be worked out only numerically. However, as far as the author could ascertain, in the temperature range usually encountered with gear steels the temperature variation of the thermal contact coefficient is small enough to warrant the original flash temperature formula without incurring too poor an approximation.

*Refinements for maldistribution of the load.*—Now comes a refinement that will more often prove necessary both in gear design and in the interpretation of test results obtained in gear testing machines or in disk machines. This is the refinement to be introduced in cases where the load is maldistributed over the width of the tooth face or disk, for instance through misalignment as caused by inaccurate assembly and/or by dis-



tortion of supporting elements like shafts or housings under load. Another cause may lie in certain manufacturing inaccuracies such as conicity of disks.

Such a maldistribution of load will result in some amount of tapering of the conjunction zone so that its width will vary in the longitudinal direction (see fig. 2). It has already been observed that flash temperatures will then no longer be uniform in the longitudinal direction and will be rather sensitive toward the local width of the conjunction zone. For further analytical details the reader is referred to Appendix B.

In conclusion, the author is under the impression that in the majority of experiments so far performed on gears or disks, corrections for the tapering of conjunction zones by maldistribution of the load should not be ignored. Fortunately, once the amount of taper is known, the correction of the flash temperature component can easily be performed (see Appendix B).

*Generalization for unequal bulk temperatures.*—Still another generalization is sometimes necessitated by too great a difference between the bulk temperatures,  $T_{b,1}$  and  $T_{b,2}$ , of the two rubbing bodies. In fact, the original flash temperature theory was specialized for the cases where the two bulk temperatures may be put equal. Accordingly, in equation (1a) for the conjunction temperature, the bulk temperature has been assigned for the common value,  $T_b = T_{b,1} = T_{b,2}$ .

But even in the general case of two unequal bulk temperatures, because the assumption about the absence of interfacial temperature jumps is retained, will the conjunction temperature be common to the two rubbing surfaces. It follows that the flash temperature component,  $T_{f,1}$ , of one surface is no longer equal to the one,  $T_{f,2}$ , of the other (see eq. (1a)). It will be shown in Appendix C that, at values of  $n = b_1\sqrt{V_1}/b_2\sqrt{V_2}$  that differ not too much from unity, i.e., in the range

$$\frac{1}{5} \leq n = \frac{b_1\sqrt{V_1}}{b_2\sqrt{V_2}} \leq 5 \quad (4a)$$

the following simple formula will hold to a good approximation:

$$T_c \cong 1.11 \frac{fW |V_1 - V_2|}{(b_1\sqrt{V_2} + b_2\sqrt{V_1}) \cdot \sqrt{w}} + \frac{1}{2}(T_{b,1} + T_{b,2}) \quad (4b)$$

The present generalization for unequal bulk temperatures is thus seen to work out in a plain manner. That is, in the original equation (1a) for equal bulk temperatures, after having therein substituted equation (1b), one may simply replace the originally common bulk temperature,  $T_b$ , by the average,  $\frac{1}{2}(T_{b,1} + T_{b,2})$  of the two unequal bulk temperatures.

*Refinement for thermal bulging in the conjunction zone.*—A refinement that is not so obvious as the above ones is that suggested and worked out by Ling and Mow (ref. 26) and also by Korovchinski (ref. 27). It relates

to the phenomenon that the temperature fields created in the two rubbing bodies by the frictional heat cause not only thermal stresses<sup>15</sup> but also thermoelastic strains in these bodies. Thus the contour of either body will tend to be altered by what may be called "thermal bulging." This effect, under the given load and as compared with the original flash temperature theory where the thermal bulging was neglected, will in turn tend to alter both the width of the conjunction zone and the distribution of the pressures. Accordingly the distribution of the frictional heat generated and that of the flash temperatures will also be altered.

All these effects will interact and thus be coupled in a rather complicated way. Korovchinski succeeded in solving this problem for cases where the ensuing distribution of conjunction pressures and that of the frictional heat have the same shape, the coefficient of friction being assumed constant throughout the band-shaped conjunction zone. Thus his solution does not apply to cases where a full elastohydrodynamic film interposes itself in the conjunction zone. Rather, Korovchinski's solution is representative of dry friction, boundary lubrication, or of partial elastohydrodynamic lubrication under conditions so severe that a sizable fraction of the total load is borne by direct contact.

Ling and Mow developed a numerical method of solution to problems that show an even more intricate interaction and coupling in that they did account for the presence of a full elastohydrodynamic film. Since in their cases the frictional heat is generated by viscous shearing throughout the film, and no longer at a common interface in the conjunction zone, the coupling of all the effects to be considered became so complicated that they had to simplify the problem. Indeed, they substituted a distribution of the frictional heat that had already been determined by other investigators for fully elastohydrodynamic cases where the thermal bulging effect was neglected altogether.

Ling and Mow's solution is not wholly representative of the states of fully elastohydrodynamic lubrication for which it was specifically set up. It may well be that for the states of rather severe partial elastohydrodynamic lubrication, which are considered most typical of incipient scoring, Korovchinski's solution comes closer to reality than Ling and Mow's. Since, moreover, the former solution has in its way been worked out more fully, and since the author aims first of all at an approximate idea about the conditions beyond which the effects of thermal bulging on the flash temperature component can no longer be neglected, he will follow up this solution of Korovchinski.

---

<sup>15</sup> With a view to their effect on pitting fatigue, the thermal stresses alone had already been considered by Kelley (ref. 37) who to this end superimposed these stresses upon the Hertzian stress field that is caused by the Hertzian semielliptic distributions of pressure and frictional shear stresses in the conjunction zone.

Korovchinski found his most important results to be expressible by means of what is here called his "thermal-bulging parameter,"  $K$ , which is a dimensionless group. In the particular case where reduced elastic modulus,  $E_r = E/(1-\nu^2)$ , Poisson's ratio,  $\nu$ , heat conductivity,  $k$ , and the coefficient of linear thermal expansion  $\epsilon$ , are the same for the two rubbing materials Korovchinski's thermal-bulging parameter may be defined by:<sup>16</sup>

$$K = \frac{1}{2\pi} \cdot \frac{fV_s \cdot W}{\sigma_H} \cdot \frac{(1+\nu) \cdot E_r \cdot \epsilon}{k} \quad (5a)$$

where, as in the foregoing,  $f$  denotes the coefficient of friction,  $V_s = |V_1 - V_2|$  the sliding velocity,  $W$  the load per unit length of the conjunction zone, and  $\sigma_H$  the maximum Hertzian pressure that would occur under conditions of elastic contact in the absence of thermal bulging.

Substituting the well-known expression for the Hertzian maximum pressure,  $K$  may be rewritten as (see Appendix D):

$$K = \frac{1}{\sqrt{2\pi}} \cdot fV_s (WR)^{1/2} \left[ \frac{\epsilon}{K} \cdot \left( \frac{1+\nu}{1-\nu} \cdot E \right)^{1/2} \right] \quad (5b)$$

where  $E$  denotes Young's modulus of elasticity, which is ordinarily used instead of the reduced modulus,  $E_r$ .<sup>17</sup>

In a fairly wide range of  $K$ , Korovchinski evaluated the maximum conjunction pressure,  $\sigma_K$ , that will obtain in the presence of the thermal bulging effect. He found the ratio between  $\sigma_K$  and  $\sigma_H$ , the Hertzian maximum pressure obtaining in the absence of the thermal bulging effect, to be a function solely of  $K$ . Within the somewhat limited range,

$$0 \leq K \leq 2 \quad (6a)$$

which, however, will seldom be succeeded in actual gear practice. Korovchinski found that the following expression could be fitted to a reasonable approximation to his original numerical results:

$$\frac{\sigma_K}{\sigma_H} \cong 1 + 0.62K \quad (6b)$$

It may be derived that at any given load and in range (6a) the width,  $w_K$ , of the conjunction area according to Korovchinski must become

<sup>16</sup> For simplification only this particular case is considered here, but this case is representative of many cases found both in actual gears and on disk machines.

<sup>17</sup> Note that in equation (5b) all the characteristics of the material have been grouped within the brackets. For steels of the kinds usual for gears, this group will show a value about equal to  $5 \times 10^{-3} \text{ kgf}^{-1/2} \text{ cm}^{-1} \text{ sec}$ , so that equation (5b) reduces to about  $K = 2 \times 10^{-3} \cdot fV_s (WR)^{1/2}$ , where  $V_s$ ,  $W$ , and  $R$  have to be expressed in the kgf-cm-sec-system of units.



narrower in about inverse proportion to  $\sigma_K$ . Accordingly, within the range concerned, one may write to about the same approximation as inherent in equation (6b):

$$\frac{w_K}{w_H} \cong (1 + 0.62K)^{-1} \quad (7)$$

where  $w_H$  denotes the Hertzian width that would obtain in the absence of thermal bulging.

Further it may be derived also that in range (6a) the effect of thermal bulging or, say, of the narrowing of the conjunction area and of the attendant increase in the conjunction pressures, on flash temperature may be expressed by including a factor  $(1 + 0.62K)^{1/2}$  in the original flash temperature equation (2a) as follows:

$$T_{f,K} \cong 0.62(1 + 0.62K)^{1/2} \cdot fW^{3/4} |V_1 - V_2| \cdot E_r^{1/4} \cdot (bR^{1/4})^{-1} \quad (8)$$

where the second suffix of  $T_{f,K}$  has been introduced to distinguish this expression<sup>18</sup> from the original equation (2b).

In actual gear practice it will not often happen that Korovchinski's thermal bulging parameter,  $K$ , will exceed 1.0 (see eq. (5b) and footnote 17). For instance, at this particular value the correction factor  $(1 + 0.62K)^{1/2}$  amounts to 1.27 so that the flash temperature would have to be corrected by plus 27 percent.

Last but not least, it has still to be considered that Korovchinski based his theory on the rather oversimplifying assumption that the temperature field, and thus the thermal bulging, in either rubbing body would be the same as in the comparatively simple cases where such a body is stationary with respect to the source of frictional heat concerned. His theory will, therefore, overestimate the amount of thermal bulging and also its effects (see eq. (8)) on flash temperature. Yet, as an overestimate, equation (8) may serve a useful purpose. For instance, whenever this formula indicates a negligible influence of thermal bulging, this may safely be assumed to be the case.

*Generalization to states of partial hydrodynamic lubrication.*—Another generalization to be treated relates to certain phenomena that will arise in states of partial lubrication and that might conceivably necessitate refinements in the flash temperature theory or, say, corrections to the original flash temperature formula. In fact, judging from various findings, both in actual practice and in testing machines, it is thought that at the loads at which incipient scoring occurs in gears, a state of rather severe

<sup>18</sup> It is merely a coincidence that the first factor, 0.62, which is common to equations (2b) and (8), has the same value as Korovchinski's coefficient 0.62 in the correction factor for thermal bulging,  $(1 + 0.62K)^{1/2}$ .

partial elastohydrodynamic lubrication will, more often than not, have already been reached.<sup>19</sup>

Let us now contemplate such states of partial elastohydrodynamic lubrication, the author considering those most typical of incipient scoring. It is assumed that these states are so severe that at least some 10 or 20 percent of the total load is borne by contact which will be dispersed more or less throughout the conjunction zone. Of course, the balance of the total load will then be borne by the elastohydrodynamic generation of lubricant pressures in what remains of the film, i.e., in the interstices remaining between the various areas of real contact.

The frictional heat will now be generated partly on the areas of real contact through boundary friction, and for the remainder in the interstices through viscous shearing of an elastohydrodynamic nature. Now, the total frictional heat may be described by  $fW |V_1 - V_2|$ , as in the original flash temperature formula and provided that  $f$  denotes an overall or average coefficient of friction. In fact, the coefficient of friction prevailing on the areas of real contact may be expected to be equal to that characteristic of the given oil under conditions of boundary lubrications so that it will amount to something like 0.15.<sup>20</sup>

Further it is known from measurements that, under operating conditions similar to those contemplated, the overall or average coefficient of friction,  $f$ , will range from about 0.05 to 0.10. It follows that the coefficient of viscous friction to be assigned to the interstices, that is, to what remains of the elastohydrodynamic film, cannot be much smaller than 0.05, and may even be somewhat greater.

Consequently it may be assumed that the shape of the distribution of the frictional heat over the width of the conjunction zone will be far from jagged. That is, its peaks, which are to occur on the sites of the areas of real contact, are not expected to be very sharp. It is admitted though that the present distributions of the frictional heat will not be as smooth as the Hertzian semielliptic distribution, the latter being the one having been assumed in the derivation of the original flash temperature equations (1b) and (2a), and for which the numerical factor, 1.11, in these same equations is valid.

For quite a variety of shapes of distributions of frictional heat, with the same resultant,  $fW |V_1 - V_2|$ , and of which several were much less smooth than Hertz's semielliptic distribution, the author has calculated

<sup>19</sup> However, the possibility is not excluded that a full fluid film might persist up to incipient scoring, so that scoring would be caused by some kind of sudden collapse of that film. This possibility will be considered also in the next section, in an attempt to conceive one or more mechanisms that might result in a collapse, perhaps through some kind of instability in the flow of lubricant and/or heat in the full fluid films concerned.

<sup>20</sup> It is thought that Terauchi and Miyao (ref. 29) overestimated this particular coefficient of friction by putting it equal to 0.5.



the corresponding maximum flash temperatures,  $T_f$ . Within the fairly wide range of shapes thus explored, he invariably found  $T_f$  to be fairly insensitive toward the shape. In fact, the numerical factors that had to be substituted, for the various shapes, in the flash temperature equation (2a) never deviated more than a few percent from the original numerical factor, 1.11.

It is thus concluded that even under conditions of partial elastohydrodynamic lubrication, provided that these are not so mild that they come very close indeed to full elastohydrodynamic lubrication, the original flash temperature equation (2a), including its numerical factor, 1.11, may be retained to a reasonable approximation. Indeed, any correction descriptive of changes in the shape of the distribution of the frictional heat is deemed unnecessary under the conditions contemplated.

*Generalization for marginal elastohydrodynamic lubrication.*—Let us now go into the question of the need for a refinement of the flash temperature theory as regard the cases of marginal elastohydrodynamic lubrication. By definition these are the borderline cases where, on an overall basis, the total load is borne practically wholly by the pressures generated hydrodynamically in the lubricant film. But this implies that only a tiny fraction of the total load, which fraction is very much smaller than in the above-discussed severe partial elastohydrodynamic lubrication, is carried by those asperities on which some amount of contact is established, usually only intermittently. Of course, this contact will not only be of a rather occasional but also of a highly disperse nature.

Cases are even conceivable where during certain intervals there will be no contact at all. Such intervals are characterized by full elastohydrodynamic lubrication, that is, full separation of the rubbing surfaces, transient as it may be. The present transient cases will be considered first, even though, due to the total absence of any contact, there will be no risk of scoring whatsoever throughout the intervals concerned. It will prove that these cases provide a useful basis to the considerations to be set up later for the more complicated cases during the intermediate intervals where contact does prevail. The latter cases are the only ones where it makes sense to go into the risk of scoring. But the contactless cases will first be treated since these are the only ones where some basic knowledge has already been gained on the temperatures of the two rubbing surfaces in their conjunction zone.

For the present purpose, studies on the thermal phenomena occurring in full elastohydrodynamic lubrication, that is, in the above-defined contactless cases, have proved increasingly useful as they were being published in the course of time (Grubin, ref. 38; Crook, ref. 39; Dowson and Whitaker, ref. 40; Cheng and Sternlicht, ref. 41; and Manton et al., ref. 42). Therefore, the author will check his derivations mainly on the basis of the latest study available, that of Manton et al.



In the first place we shall have to consider a basic difference between the original flash temperature theory and the thermal theory for full elastohydrodynamic lubrication. In the former theory the frictional heat is assumed to be generated in a planar form, i.e., by the solid or boundary friction that occurs at the interface that is considered common to the two rubbing surfaces in their conjunction zone. But in the nonisothermal elastohydrodynamic theory, this assumption can no longer be upheld since the frictional heat is here generated in a spatial form, that is, by viscous shear<sup>21</sup> throughout the lubricant film. Certain consequences of the spatial nature of the frictional heat are worthy of note for the present purpose and can be explained as follows.

Apart from the temperature field in either rubbing body, now a third one will also have to be explored, the one generated in the film. A major difficulty is that the coupling between the two temperature fields in the rubbing bodies is no longer direct, i.e., not through the absence of interfacial temperature jumps, but indirect, i.e., through the temperature field in the film. True, the assumption about the absence of interfacial temperature jumps is still valid, but it will now have to be applied twice, that is individually to either solid/liquid interface. In fact, the temperature of the lubricant at some point of the rubbing surface concerned will have to be put equal to that of the rubbing material at that same point. But the possibility cannot thus be precluded that the two distributions of surface temperatures will be different or, say, that there will be temperature differentials across the film. The reality of this possibility has been well demonstrated in the above-mentioned theoretical papers.

Let us further confine ourselves, as in the papers so far published, to the cases where the heat flow in the lubricant film, and from thereon toward either adjoining rubbing body, is not interfered with by some heat source external to the film, such as that constituted by a hot rotor that is thermally connected to one of the rubbing bodies. In such cases the maximum temperatures of the entire system cannot but lie inside the film. More precisely, the variation of the lubricant temperature in the direction of the normal will show a maximum of which both the value and location will be characteristic of the location of the pair of opposite points considered for the two rubbing surfaces. Further there will be a locus of the maximum temperatures that extends throughout the length of the film and that in general will be a curved surface lying between and nowhere touching the two rubbing surfaces.

At the above-defined locus there will, of course, be no heat flow at all, but on any given side of the locus the heat flow will become increasingly

---

<sup>21</sup> Adiabatic compression and decompression as the film pressures increase and decrease and their positive and negative contribution, respectively, may not be neglected when the maximum film pressure should attain a really high level. But this phenomenon is not so typical of the heating in lubricant films in general.

stronger until it reaches the adjacent rubbing surface. Summarizing, the heat developed in the film by viscous shear may be conceived to split up into two heat flows, both virtually emanating from the locus and having opposite directions. These two heat flows act together in carrying away the heat generated in the film toward the two rubbing surfaces. Finally it can be demonstrated that, with respect to the present "two-way" cross-conduction of the heat, one may neglect the other mechanisms conceivable for carrying away the heat, i.e., the convection of heat by the oil acting as a heat carrier while it flows from the inlet toward the outlet of the film, and the heat conduction through the oil in that same longitudinal direction.

Thus either portion of the lubricant film that lies between the locus of the maximum temperatures and the rubbing surface concerned may be assigned a thermal resistance that has to be overcome by the heat flow in the portion considered. Further it can be reasoned that the locus should lie closer to the slower than to the faster of the rubbing surfaces.<sup>22</sup> Accordingly it may be expected that the temperature distribution on the slower surface will be at a higher overall level than that on the faster surface (see ref. 42).

Let us now examine in what way the above thermal picture can be related to the risk of scoring. In this connection it has again to be emphasized that scoring can only set in at places where contact occurs between two rubbing surfaces.<sup>23</sup>

In the marginal regime considered, such contacts will indeed take place. Further their occasional and intermittent nature involves their temperatures just prior to the occurrence of incipient or "grazing" contact between two opposite points of the rubbing surfaces. The temperatures of these spots will be equal to those found for these same points from the above two temperature distributions for full elastohydrodynamic lubrication. In view of the difference in the overall levels of these distributions, the "opposite" temperatures of the two contacting points will also be different just prior to the establishment of the contact. However, from the initiation of the contact onward, and as long as it lasts, the "opposite" temperatures, on the assumption of the absence of interfacial temperature jumps at contacts, must so adapt themselves as to be equal. It is an important question how to reconcile, especially at the very

---

<sup>22</sup> This feature is indeed brought out by the numerical evaluations available in literature.

<sup>23</sup> True, the maximum temperature in the entire system will not occur on the rubbing surfaces, but inside the film. It has been shown by Crook (ref. 39), Dowson and Higginson (ref. 43) and Manton et al. (ref. 42) that it may even be more than 50° C higher than the highest temperature on the rubbing surfaces. Even so it is thought that the maximum initial contact temperature, which is investigated in what follows and which is not affected by the maximum oil temperature in the film, is decisive for the risk of scoring.



initiation of the contact, this requirement of equality with the difference that elastohydrodynamically exists just prior to that initiation.

The present question can be resolved by means of a well-known problem in the theory of heat conduction. This is the problem of stationary bodies of which the surfaces, while being kept apart, are initially at different temperatures,  $T_1$  and  $T_2$ , and at some instant are brought into contact and further held so. It is required to find out what happens, particularly at the initial instant, to the temperatures of the two bodies in their common contact area.<sup>24</sup>

The solution to the present problem may be found in several textbooks on heat conduction (see also ref. 44) and is primarily based on the assumption of the absence of a temperature jump. Although in the present case the bodies are not stationary but sliding, it may be considered that at the instant concerned their surfaces just start touching so that the frictional heat developed by solid or boundary friction at the interface is negligible yet. Therefore, at this instant the condition must still be satisfied that at the interface there is no discontinuity in the heat flow from the hotter to the cooler body.

It is thus found that at the initial instant an interfacial temperature,  $T_i$ , which, as an incipient contact temperature, is common to the two bodies, develops instantaneously. More precisely,  $T_i$  will be intermediate between the temperatures  $T_1$  and  $T_2$  shown by the rubbing surfaces just prior to contact and at the two opposite points at which contact is to be established immediately thereafter, as follows:

$$T_i = \frac{b_1 T_1 + b_2 T_2}{b_1 + b_2} \quad (9a)$$

where  $b_1$  and  $b_2$  denote the thermal contact coefficients of the rubbing materials 1 and 2, respectively, and as defined already in connection with with the flash temperature equations (2a), (2b), and (2c).

It is readily seen from equation (9a) that  $T_i$  indeed is intermediate to  $T_1$  and  $T_2$ . Accordingly the distribution of  $T_i$  over the conjunction zone will have some such intermediate position with respect to the distributions of  $T_1$  and  $T_2$  as indicated in figure 4.

The exact determination of the incipient contact temperature,  $T_i$ , follows from equation (9a), once the temperatures  $T_1$  and  $T_2$  are known. As is evident from the foregoing, the latter temperatures can be determined from the thermal theory of elastohydrodynamic lubrication, as has indeed been done by Manton et al. (ref. 42). They have restricted themselves to the particular case where the thermal coefficients,  $b_1$  and  $b_2$ ,

---

<sup>24</sup> Details about the time history after the initial instant will not be given since they will prove unnecessary.



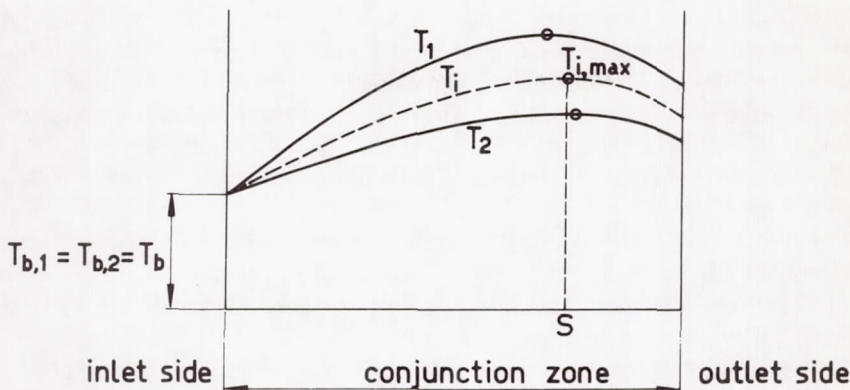


FIGURE 4.—Intermediate position of the spotwise distribution of the incipient contact temperatures,  $T_i$ .

are taken equal so that, in accordance with equation (9a),  $T_i$  reduces just to the arithmetic average of  $T_1$  and  $T_2$ , as follows:

$$T_i = \frac{1}{2} (T_1 + T_2) \quad (9b)$$

They have further restricted themselves to equality of the bulk temperatures of the two rubbing bodies or, say, of the temperatures of the two rubbing surfaces in the inlet cross-section of the film. At the inlet, therefore, both  $T_1$  and  $T_2$  equal  $T_b = T_{b,1} = T_{b,2}$  (see fig. 4).

Notwithstanding the above restrictions, the work of Manton et al. is deemed sufficiently representative of meshing teeth. Accordingly, as to the distribution of the incipient contact temperatures,  $T_i$ , the author has confined himself to studying the consequences of the various sets of distributions of  $T_1$  and  $T_2$  that were assessed numerically by Manton et al. for their various sets of elastohydrodynamic operating conditions that together covered a fairly wide range.

The main conclusion, as derived in Appendix D, is that throughout Manton's range, the maximum incipient contact temperature  $T_{i,max}$ , as derived from the elastohydrodynamically valid distributions of the surface temperatures of the two rubbing bodies may be put equal (to a reasonable approximation) to the maximum flash temperature,  $T_f$ , that follows from the original flash temperature equation (2a) if one substitutes for  $fW |V_1 - V_2|$  the heat actually developed per unit width in the fully elastohydrodynamic film.

Since this range is deemed representative of usual actual gear practice, it is thought that the above conclusion may indeed be applied to that practice. Thus, it would be applicable also to disk machines and other simulating machines in so far as their operational range is, indeed, representative of that same practice.

It is not really necessary to assess the heat developed per unit length of the fully elastohydrodynamic film from theory. Experimental knowledge of the coefficient of friction,  $f$ , will suffice in calculating the maximum incipient contact temperature from the original flash temperature equation. After all, substitution of the value of  $fW |V_1 - V_2|$  in this equation will suffice, unit load  $W$  and sliding velocity  $|V_1 - V_2|$  being also known.

Let us revert to the physical meaning of the distribution of the incipient contact temperatures,  $T_i$ , and particularly of its maximum,  $T_{i,\max}$  (see fig. 4), in terms of the risk of scoring. In the regime contemplated, that of marginal elastohydrodynamic lubrication, contact will never be established all over the conjunction zone, and thus the same will be true of the distribution of the incipient contact temperatures,  $T_i$ , as indicated in figure 4. In fact, contacts will here occur occasionally and intermittently, and at spots that will not only be dispersed over the conjunction zone but also may change their position all the time. The incipient contact temperature of each individual spot depends on the latter's location. Thus the distribution of these temperatures, such as the one indicated by the  $T_i$  curve in figure 4, has to be interpreted as a "spotwise distribution." Indeed, one may read such incipient and local contact temperatures spotwise.

Of course, the maximum incipient contact temperatures, all equal to the above-defined  $T_{i,\max}$  (see also fig. 4), will be reached only at those contact spots that happen to occur at the particular film cross-section where  $T_i$  reaches its maximum,  $T_{i,\max}$ . (See the cross-section marked by S on the horizontal axis of fig. 4.) It can be assumed that, irrespective of the regime of lubrication, the risk of scoring of two rubbing surfaces is determined by the highest temperature occurring in the conjunction zone on either incipient or permanent contact areas. Therefore, the range of applicability of the original flash temperature theory may be widened as follows.

Whenever lubrication is not permanently of the fully hydrodynamic type, the risk of scoring may be judged from the maximum flash temperature,  $T_f$ , as calculated from the original flash temperature equation with the known heat,  $fW |V_1 - V_2|$ , developed per unit length of the conjunction zone.

#### Refinements in the Theoretical Prediction of Bulk Temperatures

Inaccuracies in estimates of maximum conjunction temperatures,  $T_c$ , are not only due to uncertainties in estimates of the flash temperature components,  $T_f$ , but also to those inherent in estimates of the bulk temperatures. The truth of this observation is most readily seen from equation (1a), which is valid for the particular cases where the bulk temperatures of the two rubbing bodies are equal and thus have a common



value  $T_b$ . But it is also brought out by the more general equation (4b), which applies to the most general cases where the two bulk temperatures are different.

It is very important to note that cases are not rare where the contribution of the flash temperature component to the maximum conjunction temperature, even when this should reach the scoring level, is far from markedly dominant as compared with the contribution of the bulk temperature. It follows that in general the inaccuracies in estimates of bulk temperatures have to be counted just as heavily as, or sometimes even heavier than, those in estimates of flash temperature components.<sup>25</sup>

It is admitted that bulk temperatures may be measured much more easily than conjunction temperatures or their flash temperature components. But this is small help to the designer of gears and other machine elements having counterformal rubbing surfaces. He is rather in need of being able to reliably predict the maximum conjunction temperatures that are going to develop during the actual operation of his conceptual procreation.

As may be evidenced by the foregoing account, progress in predictive accuracy has been promising insofar as the flash temperature component is concerned. But as long ago as 1961, it occurred to the author that the predictive accuracy of the bulk temperature component of conjunction temperatures had fallen so far behind that of the flash temperature component which even at that time was in a comparatively backward stage.

True, certain sets of basic data indispensable for predicting bulk temperatures, such as those on the heat transfer from gear boxes to the ambient air and the heat generation on meshing tooth faces and by churning in oil baths, had long been available. But even now there is a paucity of sets of basic data that can be cast into the form of sufficiently strict predictive correlations.<sup>26</sup> However, another shortcoming was considered even more important, namely, the utter lack of a well rounded-off theory into which to assimilate systematically the various sets of basic data in order to achieve a reliable prediction of the bulk temperatures.

This situation prompted the author to develop the theory that he prefers to call the "thermal network theory." In 1963 the author's then student, D. Landheer, put this theory to test in two simple gear trans-

---

<sup>25</sup> In many cases uncertainties about the bulk temperature of gears have prevented the postulate about the constancy of scoring temperature to be verified dependably from actual gear practice. (See the comments on their fig. 11.14 by Kelley and Lemanski (ref. 24).)

<sup>26</sup> To make such correlations as universal as possible, they should preferably be worked out in a dimensionless form. For instance, for churning power in oil baths, this method was followed by the author (ref. 45).



missions. The latter succeeded in satisfactorily corroborating the theory through demonstrating the bulk temperatures, measured at various elements, to come fairly close to those predicted. A concise outline of the thermal network theory will be given in what follows.<sup>27</sup>

*Outline of the thermal network theory.*—The most general objective of the thermal network theory as applied to gear transmissions is to predict the bulk temperatures that will develop in at least the most important elements of such transmissions under any set of specified conditions of loading, speed, and cooling.

Elements certainly to be classed as thermally significant are those on or in which frictional heat is generated, i.e., the gears, bearings, seals, and the oil bath, if any. The reason for the thermal significance of each such frictional element lies in the fact that its contribution to the total frictional losses of the gear transmission is closely interrelated with its bulk temperature. In fact, all these contributions are affected by hydrodynamic phenomena and thus by some representative viscosity of the oil, which in turn is determined by the bulk temperature of the frictional element concerned.

For the particular purpose of predicting the maximum conjunction temperature, and thereby the risk of scoring, in meshing zones of gears, the most important single result that may be yielded by the thermal network theory is the bulk temperature representative of the pair of gears to be considered. After all, only after the bulk temperature has been predicted, as a component of the conjunction temperature sought, can the flash temperature component be superimposed. Moreover, the latter component is proportional to the coefficient of tooth friction (see eqs. (1b), (2a), (2b), or (2c)), and this depends through the viscosity representative of the meshing zone of the pair of gears concerned, again on the bulk temperature.

Now it is important to observe that in principle the bulk temperature representative of a meshing zone cannot be assessed by considering the given pair of gears in isolation from the other thermally active elements that also generate frictional heat. In fact, all of the thermally active elements, including the meshing zone, are thermally interconnected, either direct at some interface, such as gears and the oil bath, or indirect through one or more connecting elements, as is the case of gears and bearings which are connected by shafts. In such connecting elements, which are nonfrictional or, say, thermally passive elements, the thermal interconnection takes place by heat conduction.

The thermal interconnections are a complicating factor. They provide

---

<sup>27</sup> In 1968 a more elaborate outline was given by the author within the scope of a series of lectures held in Japan. The text will soon be published in the form of a Japanese translation. A free copy of the original text in English can be made available to those interested.

paths for the heat flow generated at any given source of frictional heat to interact with the heat flow generated at any other source of frictional heat. Thus the thermal interconnection between any given pair of heat sources is, in general, redundant. This is because more than one path is available between them. From this, bulk temperature can be predicted only when its interaction with all the other bulk temperatures, which are representative of the remaining heat sources, has been accounted for in some way or another.

A further complication arises from the abovementioned fact that in a gear transmission all the heat sources are of a nonconstant and even nonlinear type. In fact, in each of them the frictional heat generated depends upon a viscosity representative of the source to be considered, and thus upon the representative bulk temperature. Now, it has already been pointed out that, because of the thermal interconnection of the various heat sources, their individual bulk temperatures will be interrelated. It follows that their individual frictional heats will also be interrelated. Accounting for the nonlinearity of the relationship between viscosity and bulk temperature, the various heat sources are thus seen to be also nonlinear in that each of the frictional heats generated will depend on the corresponding bulk temperature in a nonlinear way.

Still other complications arise under thermally unsteady states that will, in general, occur even when a gear transmission is running under steady conditions of loading and speed, for instance, under the heating-up conditions after starting. For simplicity, however, the author will confine himself here to the thermally steady cases where by definition each of the bulk temperatures of the various elements has reached its equilibrium level.<sup>28</sup> Apart from rather exceptional cases, these equilibrium levels will also be the highest levels attainable. Accordingly the particular equilibrium level of the bulk temperature that is representative of a given meshing zone will usually be also most critical from the standpoint of the risk of scoring in that zone. In this connection it is worthy of note that the kind of thermal equilibrium, as here defined for gear transmissions, is not a complete one. In fact, in any meshing zone the conjunction temperatures will still vary with time, albeit in the quasi-steady sense of a cyclic variation. More precisely, therefore, the present states would have to be termed "states of equilibrium in bulk."

The various complications still remaining for the steady cases, those of "thermal equilibrium in bulk," may explain why a sufficiently systematic method has been lacking up to 1961. The author's "thermal network

---

<sup>28</sup> Practically speaking, the following treatment is thus confined to gear transmissions running under conditions of constant loading and speed and that, moreover, have been subjected to such conditions during a period sufficiently long for reaching the state of thermal equilibrium. This heating-up period may roughly amount from half an hour for smaller transmissions up to a few hours for larger ones.



method" was then inspired by certain similarities between, on the one hand, the thermal interconnections that induce interrelationships between heat flow and bulk temperatures in gear transmissions and, on the other hand, the electrical connections that induce interrelationships between electric currents and the electric potentials or voltages in electrical networks.<sup>29</sup>

*Concepts basic to thermal networks.*—To begin with, any gear transmission can be schematically represented by a thermal network, i.e., through resolving it into "nodes" and their interconnecting "branches." The "nodes" are conceived as points where heat flows are generated by some heat source, or where they are made to dissipate, and/or where the heat flows originating from some other node meet and/or split off. The "branches" represent the various thermal interconnections between the nodes.

Whenever a node should represent a source or sink, it has to be assigned a heat-generating or heat-dissipating characteristic that describes the relationship between the bulk temperature and the heat generated or dissipated, respectively, per unit time. Every node of this particular, active kind may also have a passive character in that it may serve as a point where branches meet and/or split off from or toward other nodes. An example is provided by the conjunction zone between two meshing gear teeth in some instantaneous meshing position. Such a zone is indeed a heat source that may be represented by a node to which is assigned a heat-generating characteristic while the heat splits off either way, being partitioned among two branches, the two meshing gears.

Branches, each of which invariably connects two nodes, will in general be characterized by their thermal resistance.<sup>30</sup> The concept of thermal resistance is defined, in analogy to that according to Ohm's definition of electrical resistance, by

$$R_{th} = \frac{\Delta T}{Q} \quad (10)$$

where  $\Delta T$  denotes the temperature drop across,  $Q$  the heat flow through, and  $R_{th}$  the thermal resistance of, the thermal element that serves as the thermal connection represented by the branch concerned.

Several kinds of thermal resistance will have to be considered. Perhaps the simplest kind is the one provided by solid heat conductors, such as

<sup>29</sup> Since in the course of time well-developed mathematical procedures have become available for electrical networks, and can now be found in a great many textbooks, those interested in the present thermal network theory are well advised to consult one or more of these textbooks, for instance, references 46 to 48.

<sup>30</sup> The possibility for branches having a negligible thermal resistance is left open. In the layout of certain thermal networks that are to be made representative of gear transmissions, such branches may prove convenient.



shafts and gear blanks. Once the pattern of the heat flow is known for the conductor concerned, the relationship between temperature drop and heat flow, and thus the thermal resistance, can be determined from the theory of heat conduction in solids.

A more complicated kind of thermal resistance is found at the interface common to two conformal bodies, that is, where these are pressed together, for instance, through a shrink-fit or press-fit. It is characteristic of interfacial thermal resistances that thermal contact takes place only over the comparatively small sum-total of the highly dispersed areas of real contact. All of the individual areas of real contact are very small since they tend to concentrate on the summits of the highest asperities.

Now, the heat flow from one body to the other will have to split up into as many smaller and individual heat flows as there are individual contact areas. However, each of these individual heat flows will have to force its way through the small individual contact area concerned. It is proved in the theory of heat conduction that this constrictional flow process is associated with a temperature drop that reveals itself in either body only to a comparatively small depth. This depth is not very much greater than the height of the asperities. Since each such temperature drop is proportional to the corresponding individual heat flow, a constant "thermal constriction resistance" can be assigned to each area of real contact (compare eq. (10)).

All of these individual thermal constriction resistances are arranged in shunt so that, in accordance with the rules known from the theory of electricity, the resultant "interfacial" thermal resistance of the entire interface can be assessed by putting its reciprocal equal to the sum of the reciprocals of all the individual resistances. For further particulars, as well as for correlational formulas that may serve for predicting such interfacial thermal resistances that are virtually to be interposed between any pair of contacting conformal bodies, the reader is referred to the literature on the subject (ref. 36).

Another kind of interfacial thermal resistance, which is also of a constrictional nature, is that prevailing between counterformal surfaces and to be located in their conjunction zone. In gear transmissions this kind obtains only with counterformal surfaces that, such as occurs with meshing tooth faces or rolling bearings, are rubbing on each other (i.e., where either surface moves relative to the conjunction zone). On the one hand, the derivation of the interfacial thermal resistance is simpler than the one of the above-discussed conformal surfaces, namely, in that the conjunction zone may be conceived as the one and only area of real contact. On the other hand, the motion entails two complications.

One of these complications is bound up with the fact that when the motion shows a sliding component, such as occurs with meshing tooth faces, frictional heat will be generated in the conjunction zone. Thus such

a conjunction zone is now to be conceived not only as a thermal resistance, but also as a source of frictional heat. It follows that, in contrast to the one-way heat flow at the interface of the above-discussed fitted conformal and nonrubbing surfaces, the present heat flow will be a two-way affair. In fact, frictional heat will be partitioned between the two rubbing bodies. Accordingly either body will have to be assigned a thermal constriction resistance. The second complication relates to the influence of the motion of either body on its thermal constriction resistance.

The solution to the problem of assessing the two instantaneous thermal constriction resistances,  $R_{c,1}$  and  $R_{c,2}$ , to be attributed to the band-shaped conjunction zone between two meshing tooth faces of straight spur gears at any given meshing position, has been given by the author elsewhere (see the lectures mentioned under footnote 27, and see Appendix A), and is

$$R_{c,1} = 0.478(B\sqrt{w} \cdot b_1 \sqrt{V_1})^{-1} \quad (11a)$$

and

$$R_{c,2} = 0.478(B\sqrt{w} \cdot b_2 \sqrt{V_2})^{-1} \quad (11b)$$

where  $B$  is the length of the conjunction zone or, say, the width of the tooth faces,  $w$  the instantaneous width of that zone, and  $V_1$  and  $V_2$  the instantaneous tangential velocity of tooth faces 1 and 2, respectively, and relative to the conjunction zone, while the thermal contact coefficients of the gear materials concerned are denoted by  $b_1$  and  $b_2$ .<sup>31</sup>

The effect of the variations of the various factors causing variations of the width,  $w$ , of the conjunction zone during the meshing cycle can be expressed explicitly by the following variants of equations (11a) and (11b):

$$R_{c,1} = 0.268 \frac{E_r^{1/4}}{B} \cdot (R^{1/4} \cdot W^{1/4} \cdot b_1 \sqrt{V_1})^{-1} \quad (12a)$$

and

$$R_{c,2} = 0.268 \frac{E_r^{1/4}}{B} \cdot (R^{1/4} \cdot W^{1/4} \cdot b_2 \sqrt{V_2})^{-1} \quad (12b)$$

where  $E_r$  and  $R$  denote the reduced modulus of elasticity and the radius of conformity (see eqs. (3a) and (3b), respectively), while  $W$  stands for the actual load acting in the instantaneous meshing position considered between the two meshing tooth faces per unit width (see Appendix A for the derivation).

The concept of the instantaneous thermal constriction resistances,  $R_{c,1}$

<sup>31</sup> The derivation of equations (11a) and (11b) has been worked out on the basis of the flash temperature theory in the simplified form introduced by the author (ref. 1). Accordingly these formulas are valid in the same range of sufficiently high Péclet numbers,  $V_1 w / a_1$  and  $V_2 w / a_2$ , as the flash temperature equations (2a), (2b), and (2c). (See also footnotes 12 and 13.)



and  $R_{c,2}$ , which are to be attributed to the conjunction zone of any pair of meshing tooth faces in any of their consecutive meshing positions, has proved its worth in the following problem that is typical of the thermal network theory as applying to gear transmissions.

Basically both the existence and the variation of either thermal constriction resistance,  $R_{c,1}$  or  $R_{c,2}$ , is bound up with the phenomenon of "flash heating" which lies at the very root of the flash temperature theory. This phenomenon reveals itself in the occurrence of a "thermal skin effect" which is typical of any body over which a heat source, such as the present conjunction zone, is moving comparatively fast. Under such conditions flash temperatures will, of course, show their maximum somewhere inside the conjunction zone, that is, on the surface of the body concerned.

Now, flash temperatures will occur also at any depth below the surface. However, they will decay comparatively rapidly with increasing depth below that surface. In fact, these subsurface flash temperatures will not exceed 10 percent of the maximum flash temperature beyond depths,  $y$ , indicated by the range

$$\frac{y}{w} \geq 4.28 \left( \frac{Vw}{a} \right)^{-1/2} \quad (13)$$

where  $a$  denotes the thermal diffusivity of the rubbing material concerned, being defined by  $a = k/\rho c$ ,  $k$  is the heat conductivity,  $\rho c$  the specific heat per unit volume, and velocity  $V$  has been assumed uniform.

With velocities,  $V$ , and widths of the conjunction zone,  $w$ , representative of meshing tooth faces when these are liable to the risk of scoring, and with thermal diffusivities,  $a$ , that are found with the usual gear steels, the critical depth that constitutes the lower bound to range (13), and thus provides a measure of the thickness of the thermal skin, proves to be even smaller than the width of the conjunction zone.<sup>32</sup>

Now, the flash heating involves that the entire flash temperature field, being entrained by its source, the conjunction zone, sweeps through the entire body. Thus, although throughout any single meshing cycle the bulk temperature component may be considered to remain constant at any given fixed point of the body, the flash temperature component will vary and go through some maximum. This maximum temperature is dependent on the depth of the point. But at points underneath the thermal skin, i.e., in the bulk of the body, the variations of the flash temperature

<sup>32</sup> The derivation of range (13) has been given elsewhere and is valid for sufficiently high values of the Péclet number  $Vw/a$  (see the lectures referred to under footnote 27). Further it may be noted that the usual gear steels show thermal diffusivities that are of the order of magnitude of  $a = 0.125 \text{ cm}^2/\text{sec}$ . For instance, with this  $a$  value and at  $V = 200 \text{ cm/sec} = 78.7 \text{ in./sec} = 394 \text{ ft/min}$ , and  $w = 0.1 \text{ mm} = 0.01 \text{ cm} = 0.0039 \text{ in.}$ , the fraction  $y/w$ , as the lower bound to range (13), will amount to only 0.34.



component or, say, the transient effects of the flash heating will be barely discernible. Therefore, at any point in the bulk of the gear teeth and the gear blank, and throughout any one meshing cycle, the temperature will be essentially constant and identical to the bulk temperature. In the bulk of the gear material, in contrast to the thermal skin, the heat flow will also be essentially constant and uniformly distributed, having been evened out during its passage through that skin.

It has already been shown that the relationship between the bulk temperature and the heat flow through the bulk of the gear material can be described by means of a constant thermal resistance which can be termed the "thermal bulk resistance." Because of the extreme thinness of the thermal skins on meshing teeth, this constant resistance may, if only to a first approximation, be put equal to that of a cylindrical body having a smooth circumference, without any teeth at all. The cylindrical body will have a diameter equal to the pitch diameter of the spur gear concerned.

For this "thermal bulk resistance," it does not matter that the meshing zone, which includes the conjunction zone as the source of the heat and which in an overall sense is thus a heating zone over part of the circumference of the gear, rotates with respect to the gear. This is due to the comparatively great thermal inertia of the bulk of the gear. Put more explicitly, the time needed for one revolution of the gear, and thus for one relative revolution of the meshing zone, is negligibly small in comparison with the time needed for any discernible change in the bulk temperature of the gear.

However, the above-defined first approximation to the thermal bulk resistance of gears still needs a correction which, in many cases, proves to be far from negligible. In fact, the presence of the thermal skin, which interposes itself between the instantaneous conjunction zone and the bulk of the gear, also has to be accounted for insofar as the relationship between the heat flow and the bulk temperature is concerned. This is due to the feature that, for penetrating the thermal skin, the heat flow has to overcome constriction resistance as defined for any given meshing position by equations (11a) and (11b) or (12a) and (12b).

Now, the thermal constriction resistance, like the heat flow that penetrates the conjunction zone, will vary throughout any given meshing cycle. Moreover, one has to account for the simultaneous presence of two conjunction zones in shunt during double-tooth contact as an interval of the meshing cycle. However, by virtue of the above-mentioned thermal inertia of the bulk of the gear, for the present purpose one may reckon with a suitable average,  $R_{c,av}$ , of the thermal constriction resistance over one complete meshing cycle, and also with the average heat flow penetrating from the meshing zone into the gear concerned.

It is now seen that the thermal effect of any given gear on its own bulk

temperature, and on the bulk temperatures of the other elements of the gear transmission, can be described in terms of a total thermal resistance that is obtained as the sum of the thermal bulk resistance and the above-mentioned average thermal constriction resistance of the gear concerned.

Finally, apart from the above-treated interfacial thermal resistances originating at solid/solid interfaces, other kinds have to be considered. These are the thermal resistances offered to heat flows that have to penetrate a solid/liquid, solid/gas, or liquid/gas interface.

In gear transmissions solid/liquid interfaces may be exemplified by those where thermal contact occurs between a gear immersed in an oil bath and by those between the inner wall of the casing and an oil bath. An example of a solid/gas interface is that between the outer wall of the casing and the ambient air.

For both these kinds of interfaces, the thermal resistance is related to the so-called heat transfer or film coefficient,  $h$ , which is a constant of proportionality in the following relationship of Newton:

$$Q = hA \Delta T \quad (14)$$

where  $Q$  is the heat flow crossing the interface,  $A$  the area of, and  $\Delta T$  the temperature drop across, this interface.

In view of equation (10), it follows that the interfacial thermal resistance of the present kinds of interfaces can be expressed as follows in terms of the film coefficient,  $h$ .<sup>33</sup>

$$R_{th} = \frac{1}{hA} \quad (15)$$

The various concepts discussed in the foregoing have put us in a position to apply the thermal network theory to a variety of gear transmissions. To this end the gear transmission to be considered has first to be resolved into a representative thermal network.

The pattern of such a network must, of course, correspond with the layout of the various elements of the gear transmission that are basic to the thermal characteristics and phenomena to be accounted for. This involves that the various sources of frictional heat have to be represented by "active nodes," as defined above. To all of these active nodes a generating characteristic has to be assigned that describes the heat flow generated by the frictional heat as a function of the representative viscosity and, thus, for the given oil, as a function of the bulk temperature that is representative of the heat source concerned.

Further the thermal connections between all of the heat sources and

<sup>33</sup> For data on the heat transfer on solid/liquid and solid/gas interfaces in gear transmissions see, among other papers, the works of Wellauer (ref.49) and Ohlendorf (ref. 50).



the heat sinks, such as the ambient air, will have to be represented by branches of the network to each of which a thermal resistance has to be assigned. For every heat source that is coincident with a conjunction zone between two counterformal rubbing bodies, like the meshing tooth faces, the representative active node has to be provided with a suitably averaged thermal constriction resistance.

*Illustrative application of the thermal network theory.*—To illustrate the application of the major techniques of the thermal network theory, the author will now review the results obtained in 1963 for one of the two gear transmissions investigated by the author's then student, D. Landheer. The gear transmission considered is the test gear box used by Ohlendorf (ref. 50) in his systematic tests on frictional losses and heating phenomena. This gear box is depicted schematically in figure 5.

Even though the data obtainable from Ohlendorf's experimental work are much more extensive than those usually available for design calculations of gear transmissions used in actual practice, some of the basic data needed for applying the thermal network theory, such as certain kinds of thermal resistances and heat-generating characteristics, could be estimated only in a rather roundabout way.

The topography of the thermal network set up by Landheer as sufficiently representative of Ohlendorf's gear box is reproduced in figure 6. Because of the symmetry of this gear box with respect to its meshing zone (see M in fig. 5), it sufficed to represent only one-half of it. Indeed, figure 6 represents the righthand half of the gear box, which contains only one of the two gears and two of the four rolling bearings. These two bearings are marked by B3 and B4, in figure 5 as well as in figure 6, to facilitate also the identification of various other elements in the thermal network of figure 6, and have been coded by the same numbers or letters as in figure 5.

Because of the present symmetry, which involves symmetry of bulk temperatures of elements symmetrically placed in the two halves of Ohlendorf's gear box, these two halves may be assumed to be perfectly heat-insulated from each other. In other words, each of the thermal interconnections between the two halves across the plane of symmetry is assigned an infinite thermal resistance. Accordingly one may just as well thermally disconnect the two halves totally, as has been done in figure 6 by dismissing all those branches that in reality exist but that are ineffective in that each of them has an infinite thermal resistance and thus presents a complete thermal blocking.

Since the present symmetry applies also to the frictional heat,  $N_M$ , generated in the meshing zone, only one half,  $\frac{1}{2}N_M$ , has to be accounted for with the righthand half of the present gear box. This is why under the running conditions considered in the thermal network of figure 6, which were characterized by a frictional heat generation of  $N_M=250$  watt in



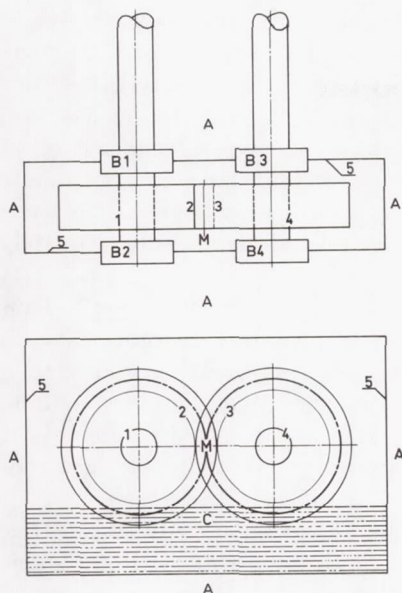


FIGURE 5.—Schematic diagram of the test gear box of Ohlendorf (ref. 50).

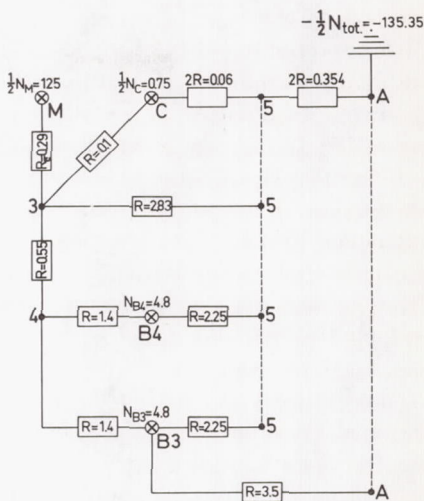


FIGURE 6.—The thermal-network representation by Landheer of the righthand half of Ohlendorf's test gear box as schematized in figure 5.

the meshing zone, only  $\frac{1}{2}N_M = 125$  watt was assigned to the active node. This active node represents this meshing zone and has been coded by  $M$  (see the upper lefthand corner of the network in fig. 6).

After the various thermal heat resistances,<sup>34</sup> the heat-generation characteristics of the various heat sources, as well as the topography of the entire thermal network have been determined, a set of simultaneous equations may be established in which the various bulk temperatures appear as the unknowns.<sup>35</sup> This can be done by systematically applying the two well-known laws of Kirchhoff as adapted to the thermal nature of the present type of network.

<sup>34</sup> The values of the various thermal resistances in the thermal network have been expressed in terms of the unit,  $^{\circ}\text{C}/\text{watt}$  (for the dimension of such resistances, see eq. (10)). For instance, the resistance  $R_c = 0.29$  watt represents the averaged thermal constriction resistance through which the meshing zone is connected to the one gear and all the other elements of the righthand half of Ohlendorf's gear box, and thus to the ultimate heat sink provided by the ambient air (see the upper righthand corner coded by  $A$  in figure 6 where half of the total frictional losses of the gear box is dissipated; that is,  $\frac{1}{2} \times 270.7 = 135.35$  watt).

<sup>35</sup> Note that, in order to obtain a neat array of the various elements thermally connected to the casing, the latter has been represented by four points, all of which have been coded by the number 5 (see also fig. 5). To indicate that, as a simplifying assumption, a uniform bulk temperature has been assigned to the casing, all these four points have been connected by what, in analogy to common usage in electrical networks, may be termed a "main bar" (see the corresponding chain-dotted line in fig. 6). The ambient air  $A$ , the ultimate heat sink, has also been indicated by such a bar.

As with electrical networks, for computational purposes the set of equations can best be represented by a matrix.<sup>36</sup> Because of the above-discussed nonlinearity of the various heat-generating characteristics, some difficulties are involved in the evaluation of the matrix. But, as has been shown by Landheer, these difficulties can be overcome by means of an iteration technique in which trial sets of bulk temperatures, especially of those representative of the various heat sources and thus of the heat-generating characteristics, are iteratively refined. The result ultimately achieved by Landheer is depicted in figure 7.

A three-dimensional representation has been chosen in figure 7. Thus the thermal network of figure 6 could be depicted in the base plane, in which the heat flows, both as to magnitude (in watts) and direction, between the various nodes have also been indicated. The various bulk temperature rises, with respect to the temperature of the ambient air as a zero reference level, have been indicated by the vertical straight lines at the corresponding nodes.

In this connection, it should be noted that the highest temperature rise indicated in figure 7,  $T_M = 97.3^\circ \text{C}$ , as a bulk temperature rise, let alone as a conjunction temperature, is merely fictitious. It has been obtained by superimposing upon the bulk temperature rise,  $T_s$ , of the gear concerned (see also figs. 4 and 5), another fictitious temperature rise that was obtained as the product of the average heat flow (125 watt; see the base plane in fig. 7) and the average thermal constriction resistance between the meshing zone and the bulk of the same gear.

The most important temperature obtaining in the meshing zone is, of course, the one indicative of the scoring risk, that is, maximum conjunction temperature that occurs during the meshing cycle. This particular temperature can be assessed by superimposing (see eq. (1)) the maximum flash temperature component, as specified by equations (2a), (2b), or (2c), upon the bulk temperature component. The latter component may be put equal to the corresponding bulk temperature which is equal to the sum of the aforementioned bulk temperature rise  $T_s$  and the zero reference temperature, i.e., the temperature of the ambient air.<sup>37</sup>

In 1963 another sample calculation on the basis of the thermal network was made by Landheer for a small industrial gear transmission which, at a pinion speed of 1460 rpm, was rated by its manufacturer at 15 hp

<sup>36</sup> For techniques that have proved useful to arrive at matrices that may be evaluated in the easiest possible manner, the reader is referred to the various textbooks on electrical networks (see, for instance, ref. 46 to 48).

<sup>37</sup> Since the bulk temperature representative of the meshing zone is thus known, the representative viscosity of the lubricating oil can simply be determined as the viscosity at that bulk temperature and under atmospheric pressure. By substituting this representative viscosity into the correlational formula of Kelley and Lemanski (ref. 24), an estimate of the coefficients of friction, including their variation during any given meshing cycle, can be obtained (for other such correlational formulas, see footnote 27).

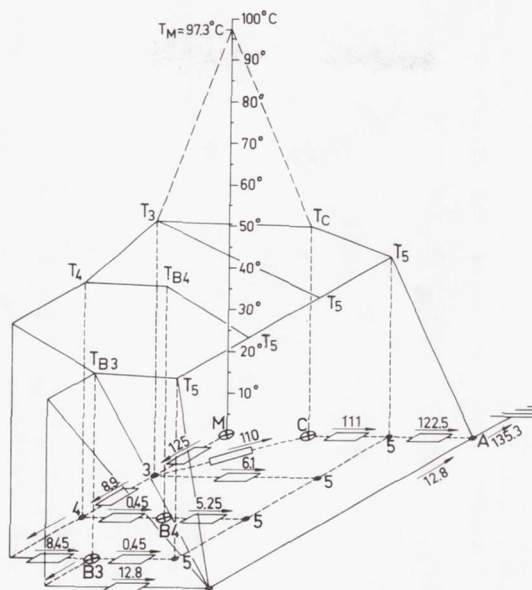


FIGURE 7.—Three-dimensional representation of the bulk temperatures and heat flows in the thermal network representing Ohlendorf's test gear box.

in continuous operation. This transmission, having a gear ratio of 96/23, was more complicated in that there was no such thermal symmetry as in Ohlendorf's test gear box. Since there was an even greater lack of reliable basic data on heat transfer and heat-generating characteristics required in applying the thermal network theory, it proved more difficult and time-consuming to arrive at sufficiently accurate estimates of such data. Yet, even with the present gear transmission, the corroboration of the theoretically predicted bulk temperatures of its various elements proved to bear promise for a wider application of the thermal network theory.

Notwithstanding this fair degree of success, a really dependable application to all kinds of gear transmissions, even though for the time being only for those operating under steady conditions of loading and speed, may well have to await stricter correlations of the aforementioned basic data than now available. This situation is somewhat comparable to that regarding the theory of hydrodynamic lubrication when it was established by Reynolds in 1886. In fact, for a long time the application of Reynold's theory was, despite its scientific and engineering soundness, severely hampered by the fact that viscosities expressed in rational units were not yet so readily available to designers and, once obtained, not so readily interpreted by them.



## SCORING MECHANISMS

## Introduction

Insofar as nonadditive mineral oils are concerned, to which the author will here confine himself, it would appear that the criterion of the constancy of the maximum conjunction temperature at incipient scoring is the only one for which a fully worked out method of design calculations, i.e., the flash temperature theory, is available for the design of gears.<sup>38</sup> But it is admitted that the present method is not totally devoid of an empirical character, no mechanism or group of mechanisms having so far been found that would explain the constancy concerned satisfactorily. The following review of scoring mechanisms that have been proposed by others, and to which a few of the author's have been added, may serve to stimulate discussion.

Designers, in general, are not so much in need of the explanation sought. Many of them will feel satisfied even with much more empirical design calculations, as long as these prove their worth in practical applications. A rational explanation is much more important to oil technologists and metallurgists since to them it may serve as a fruitful guideline in developing mineral or synthetic oils and rubbing materials that in conjunction show a better scoring performance.

One of the major difficulties in conceiving scoring mechanisms that might specifically account for the constancy of scoring temperature is that there is a paucity of significant clues about this constancy. This is true even when confining ourselves to the present, rather narrow field of the scoring performance of nonadditive mineral oils used as lubricants for the kinds of steel usually encountered in gears. Moreover, most of these clues, now to be exemplified, would appear not to be sufficiently enlightening for the present purpose.

One kind of clues might be sought in the definition of scoring. Unfortunately there is no universally accepted definition. Even as has been pointed out by Kelley and Lemanski (ref. 24), confusion has arisen because of the diversity of definitions in current usage, many not being sufficiently discriminative. This situation may explain why in certain investigations that aimed at studying scoring loads in a wide range of sliding velocities, all of the failures assessed have been indiscriminately attributed, throughout that range, to scoring. For instance, in disk tests one may well find a relationship between load at failure and sliding velocity as schematized in figure 8.

Once having plotted the experimental points, yielding an orderly pattern as in figure 8, one might be tempted to draw one single curve (see

---

<sup>38</sup> For basic data on the kinematics and conjunction of bevel and hypoid gears, see Dusev, ref. 51; Powell, ref. 52; Coleman, ref. 53; Fletcher, ref. 54; Shurygin, ref. 55; Keck, ref. 56; and Dyson, ref. 57.

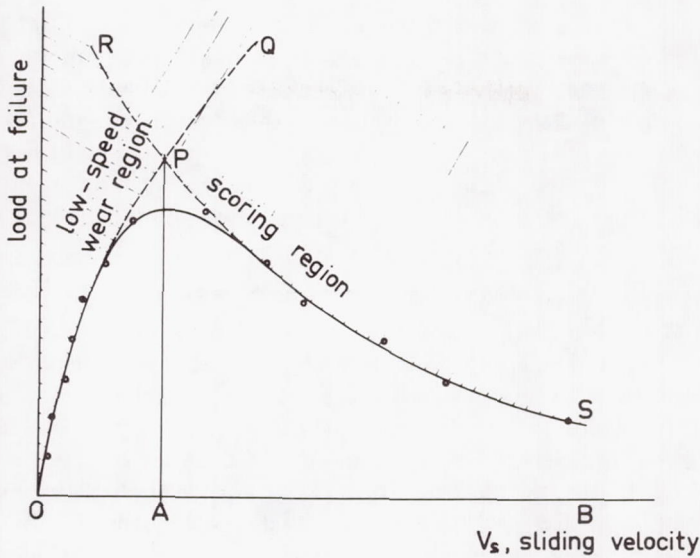


FIGURE 8.—Scoring superseded at low sliding velocities by low-speed wear.

the solid curve having a maximum in fig. 8) and conceive this to depict a relationship between scoring load and sliding velocity. But it is then overlooked that on the sole information yielded by the position of the various experimental points, while ignoring the possibility of the occurrence of different kinds of wear in the various ranges of sliding velocity, the only conclusion permissible is that the curve represents a demarcation between two regimes. The upper regime is liable to fail and the other will be free from failure. Utilizing the additional information available from visual inspection of the test specimens after failure will prove that wear of the scoring type will occur only in the regime of comparatively high sliding velocities, i.e., to the right of the maximum of the curve in figure 3. However, the wear occurring in the regime of comparatively low sliding velocities, i.e., to the left of the maximum concerned, seems to be similar to what in gear practice is often called "low-speed wear," or sometimes "abrasion."

Now, it is to be recognized that low-speed wear, in contrast to scoring, is not primarily influenced by conjunction temperatures. Therefore, to avoid confusion one should bring out the additional information available through drawing two intersecting curves (see the chain-dotted curves in fig. 8), instead of the original, single one. The curves OPQ and RPS in figure 8 will differentiate between low-speed wear and scoring, respectively. For instance, in the regime of comparatively low sliding velocities

(see regime O A in fig. 8), scoring may thus be demonstrated to be superseded by low-speed wear. In other words, in this regime, low-speed wear sets in before conjunction temperatures have become sufficiently high to cause scoring (see the mutual position of the portions OP and RP of the curves OPQ and RPS, respectively). It is thus erroneous to base verifications of the postulate about the constancy of scoring temperature upon the entire relationship between failure load and sliding velocity, not excluding the regime of low sliding velocities.

#### Clues for Conceiving Scoring Mechanisms

For conceiving scoring mechanisms, two kinds of clues may be distinguished. One kind may be sought among the features that are specified in definitions of scoring so as to distinguish this from other types of wear. Unfortunately there is quite a variety of such definitions and most of them are rather loosely termed, while none has so far been adopted universally. At least for the time being, the author prefers the following definition issued by the Institution of Mechanical Engineers: "Scoring is defined as the gross damage characterized by the formation of local welds between sliding surfaces" (ref. 58).

The clue derivable from this definition lies in the interlinkage between the feature of the formation of local welds (also called "junctions") and that of the gross damage of sliding surfaces. For instance, since the formation of local welds need not invariably result in gross damage, it is not a feature that all by itself will sufficiently discriminate between scoring and other types of wear.

In fact, it is well known from the classical work of the late F. P. Bowden that the formation of local welds may result in several types of wear, including what has here been called "low-speed wear." More precisely, it may be assumed that with metals, including gear steels, such a formation will invariably occur whenever and wherever there is contact between the virgin metals. This involves that welds are possible only when the naturally protective oxide layer on either surface breaks down, even if only in the form of local punctures. When, overlying the oxide layer, additional protective layers are present, such as a hydrodynamic and/or a boundary lubricant film, these will in general have to be punctured first before the same can happen to the oxide layer so that welds can actually be formed.

It is thus seen that the occurrence of the welds is subject to the above-mentioned conditions. But although the fulfillment of these conditions does ensure the occurrence of wear, this is not necessarily of a kind that leads to gross damage and which thus might appropriately be called "scoring." In fact, in order that the tearing brought about inevitably by sliding will result in gross damage, an additional condition has to be fulfilled. The tearing should not take place at the interface of the welds



but instead in one of the two adjacent materials. As is known from Bowden's work, this condition means that the welds should be stronger than at least one material. The kind of wear then occurring will thus be characterized by metal transfer.

It is known from general experience that gears that are liable to the risk of scoring are usually those so highly loaded that they have to be designed at least from the standpoint of the pitting risk. Therefore, these gears are usually made either of through-hardened steels or provided with a hard reinforcing layer formed in situ by some such process as carburizing or nitriding.

Further, from gear practice with such steels or reinforcing layers, gross damage and metal transfer are known to go hand in hand. The above definition of scoring thus results in the clue that, at least in the practice concerned, scoring will distinguish itself from other types of wear, such as low-speed wear, by metal transfer.

With low-speed wear, however, the absence of metal transfer should not be taken to imply that no welds are formed at all. It simply implies that the welds are comparatively weak in the above-discussed sense. Thus they cannot cause any gross damage, at least not at the fast rate that is typical of scoring. This is true despite the fact that the constituents of nonadditive mineral oils are far from outstanding in weakening the welds by contamination.

The explanation may be sought in the fact that at moderate interfacial or conjunction temperatures or at small sliding velocities, welding tendency is still comparatively small. As a result, even small degrees of contamination may make the welds the weakest link in the system. On the other hand, at the comparatively high sliding velocities that are characteristic of the scoring regime, the correspondingly higher conjunction temperatures may well promote the welding tendency to a level where the rather poor contaminating or weakening capability of nonadditive mineral oils can no longer offset that tendency.<sup>39</sup>

It would appear that the above explanation serves well also in cases such as depicted in figure 8 where, depending on the sliding velocity, two different kinds of wear may occur at one and the same load. In fact, at any given load lower than a certain maximum (see A P in fig. 8), failure may set in at two different sliding velocities. At the lower sliding velocity the conjunction temperature will be comparatively low so that there is no question of scoring and low-speed wear will occur. Only at the

---

<sup>39</sup> Beane and Lawler (ref. 59) demonstrated on their gear tester that when non-additive mineral oils were exposed in bulk to temperatures above about 205 °C (about 400 °F), the anti-scoring performance went through a minimum and at higher bulk temperatures increased. It would appear that at such high bulk temperatures oxidized constituents are formed and that their presence more than offsets the increase in welding tendency. See also Vinogradov et al. (refs. 60 and 61).

higher sliding velocity will the conjunction temperature be sufficiently high for scoring to occur.<sup>40</sup>

Clues of another kind, which may serve both for conceiving and verifying scoring mechanisms or combinations of such mechanisms, are the ones that follow from practical knowledge about the effect of various factors on the scoring performance of nonadditive mineral oils in gears. To examine the significance of such mechanisms in the verification of the postulate about the constancy of scoring temperature, which is to be carried out later on, the scoring performance will be expressed in terms of scoring temperature, i.e., the conjunction temperature at incipient scoring.

For a really general investigation into the factors that might conceivably affect scoring temperature, one should not confine oneself to the factors that, in the form of mechanical or thermal characteristics, are individually inherent in any one of the two rubbing materials or in the non-additive mineral oil. One should leave open the possibility that factors may occur that are influenced by the entire combination of these three elements. A case in point is provided by the possibility that surface activity as bound up with boundary lubrication characteristics, might also play a part in scoring performance in gears, as has indeed been suggested by some investigators (refs. 62 and 63).

In this connection it should be noted that in cases where a full fluid film completely separates the two rubbing surfaces, there is a multiple protection against scoring. In fact, when such a film should break down so that contact is established, there will still be two "stand-by" protective layers on either surface, one being the natural oxide layer and the other the boundary layer. Even during full fluid-film lubrication, a boundary layer forming a transition between the film and the natural oxide layer may be assumed to exist. This is because such a film will not exhibit any slip with respect to either rubbing surface. So at least a monomolecular layer of oil molecules must be attached sufficiently firmly to either rubbing surface. Now, such an attachment, which will take place by the kind of surface activity known as "adsorption," is typical of a boundary layer.

As a start, let us consider the survey chart published previously and presented in figure 1. In this figure, scoring temperatures,  $T_s$ , have been plotted for a variety of nonadditive mineral oils as investigated by various investigators in gear testers or disk machines.<sup>41</sup> It may be assumed

---

<sup>40</sup> If the postulate about the constancy of scoring temperature should hold good for a series of tests that results in a diagram like the one depicted in figure 8, the scoring curve P S, in contrast to the low-speed wear curve O P, would be an isotherm for conjunction temperature.

<sup>41</sup> For a more elaborate description of the derivation of this chart, including the kinds of measurements and the kind of calculations utilized in establishing it, the reader is referred to a previous paper of the author (ref. 21).



that, for any given combination of two steels used for two mating rubbing surfaces, the scoring temperature will depend on certain characteristics. Among these are the composition of the particular nonadditive mineral oil used as a lubricant. However, since it is not yet known what characteristics of such an oil influence the scoring temperature, the author for the time being, has contented himself with taking the viscosity grade<sup>42</sup> of the oil concerned as an index of its composition. True, this index may well prove too rough for this purpose, but it shows at least the merit of being readily available.

Further, in figure 1 the author has confined himself to scoring temperatures for combinations of nonadditive mineral oils of 12 different viscosity grades with steels of the kinds, and particularly of the hardness, that are usual for highly loaded gears. For each such combination the uncertainty range for each of the 12 corresponding scoring temperatures has been indicated.

Despite the ranges of uncertainty involved, an overall trend is discernible. That is, statistically speaking, with increasing viscosity grade the scoring temperature tends to increase also. But it remains to be seen whether, once the uncertainty ranges should have been narrowed by more accurate derivations of the scoring temperatures, the correlation then obtained would be much stricter than the one in figure 1. It is expected though that the overall trend under discussion would be maintained.

Nevertheless, another index of the composition of nonadditive mineral oils, one that would be more representative of the characteristics that really influence the scoring temperatures, would result in a stricter correlation. Volatility as, for instance, measured by the ASTM distillation curve of the oil concerned, might perhaps provide such a refined index. There are indications, albeit so far somewhat vague, that of two oils having the same viscosity grade and blended from the same crude, the one constituting the narrower fraction, i.e., having fewer volatile constituents, will show a somewhat higher scoring temperature. If this observation should be confirmed, the effect of volatility might prove helpful as a clue for the conception of scoring mechanisms.

Another clue, relating to the effects of the kind of rubbing materials, might perhaps be found in the fact that with steels that are appreciably softer than those included in figure 1, the scoring temperatures tend to be distinctly lower. This fact is illustrated in a chart recently set up by O'Donoghue et al. (ref. 64) along the lines of figure 1. A similar finding had already been published by Matveevsky (ref. 65) in his survey of data on scoring temperatures. Matveevsky suggests that, with the com-

---

<sup>42</sup> As usual, viscosity grade has been expressed in figure 1 as the viscosity at a standard reference temperature common to all oils. In the present chart this temperature has been taken equal to 60° C = 140° F. Of course, viscosity grade is not to be taken as a hydrodynamic factor.



paratively soft steels concerned, plastic deformation will no longer be confined to the very summits of the contacting asperities but will take place over a significant portion of the contact area and also spread deeper into the underlying metal. Thus the failure of the lubricant might be due to the joint influence of the conjunction temperature and the plastic deformation. Accordingly, scoring, as characterized by gross damage through metal transfer, might well take place at temperatures lower than those for the harder steels.

Finally other clues relating to the rubbing materials may be based on certain findings about the adverse effects of austenite as a component in the structure of the steel. In the first place, it has long been known that stainless 18/8 steel is poor as to scoring performance so that it should show a comparatively low scoring temperature (see, among others, the paper by O'Donoghue et al., ref. 64). Further it has been observed by Niemann and Assmann (ref. 66) that retained austenite in otherwise normal gear steels may also have such an adverse effect.<sup>43</sup>

#### Conception of Scoring Mechanisms

In the above-discussed definition of scoring, the one issued by the Institution of Mechanical Engineers, attention is focused on the failure of the rubbing materials, particularly as to gross damage, i.e., the result of severely overstressing their surface layer. However, for conceiving scoring mechanisms one should rather concentrate upon the basic cause of the severe overstressing, i.e., on the fact that the welds form in such a way that they constitute the strongest link in the system so that the inevitable tearing cannot but take place in the bulk material and the gross damage will be characterized by metal transfer.

Further it has to be considered that the formation of welds, even of those of the very weakest kind, can take place only when all of the protective layers have broken down, at least in the form of more or less localized punctures, so that the two virgin metals may get into contact. In cases where hydrodynamic lubrication is only partial, contact will occur by definition but before this may assume the form of virgin contact, two protective layers, the boundary layer and the oxide layer, must both be overstressed. In cases where fully hydrodynamic lubrication prevails, the

---

<sup>43</sup> Evans (ref. 67) suggested that scoring would occur when the surface temperatures are high enough to produce a transformation of steel to the austenite form, and thus should be at least about 750° C. It is thought, however, that such very high surface temperatures are not necessary at all to induce scoring, although it is admitted that during progressive scoring temperatures may well increase continually until such a high temperature level is reached and the transformation to austenite will actually set in. Thus it would appear that Evans considered an effect rather than a cause of scoring. Further it should be noted that in the above text, in contrast to Evans's steels, we are considering steels that already contain austenite before scoring sets in.

breakdown of a third protective layer, the full fluid film itself, must even precede that of the two aforementioned layers.

The overstressing and, therefore, the breakdown of each of the three protective layers may be of a mechanical and/or thermal nature. As will be shown later on, in the case of the oxide layer it may also be partly of a chemical nature since the rate of self-healing by oxidation may tend to offset the rate of breakdown.

It is thus seen that in scoring mechanisms and protective functions, the lubricating oil may in general fulfill a multifunctional role. In fact, in the form of a full fluid film it may significantly reduce both the mechanical and the thermal overstressing of the bulk material, as well as of its two superimposed protective layers, the adsorbed boundary layer and the oxide layer. Even in the form of a partially hydrodynamic film, it may still fulfill these two functions, albeit much less satisfactorily. But then still other functions come into play, the oil acting as a carrier of several kinds of constituents such as those that may contaminate and thereby weaken the welds, those that in being surface-active may aid in the self-healing of the boundary layer, and oxygen or oxygen-releasing constituents that may increase the rate of self-healing of the oxide layer.

Let us now consider various mechanisms that in the course of time have been suggested by other investigators, and including a few of the author, for the failure in the above sense of nonadditive mineral oils. The multiplicity of these mechanisms is indicative of the fact that no single mechanism or combination of mechanisms has so far been recognized as the decisive one, let alone as the one that could convincingly explain the constancy of the scoring temperature that has been postulated by the author for the oils concerned.

In the following discussion a distinction will be made between two categories of scoring mechanisms, namely:

(1) The category of scoring mechanisms primarily attributed to loss of the protective capacity of the oil through an adverse change in some characteristic of the oil as a material. For instance, the oil might somehow lose its capacity to sufficiently weaken the welds through contamination and/or its capacity to prevent the conjunction temperature becoming impermissibly high. Further a change of the kind concerned might cause the film to collapse as a structural element.

(2) The category of scoring mechanisms primarily attributed to loss of the protective capacity of the hydrodynamic oil film as a structural element. For instance, a full fluid film might somehow collapse through some kind of instability that might develop in that structural element without any irreversible change in the characteristics of the structural material, the oil.

(3) The category of scoring mechanisms that relate to loss of the pro-



tective capacity of boundary and oxide layers superimposed on the rubbing surfaces.

#### Critical Changes in Oil Characteristics

In reviewing mechanisms classifiable under the first category, it may be considered first of all that oils, being in the liquid state, will not structurally break down under compression, however high the pressure may be.<sup>44</sup>

But for the liquid state there are other possibilities of breakdown. One of these possibilities is that beyond some critical shear stress, or some critical rate of shear, the viscosity of an oil will drop gradually with any further increase in shear stress or rate of shear.

A second possibility, which seems to have originally been suggested by Borsoff (ref. 15), is that the oil might display viscoelastic characteristics, owing to the high rates of variations in pressure and shear rates to which the oil is exposed during its flow through the film, particularly in fast-running gears. Later on, Borsoff and Godet (ref. 68) have followed up this thought by introducing the concept of exposure time, defined as the time during which the oil particles in flowing through the Hertzian conjunction zone are exposed to the severe conditions prevailing in that zone. Indeed, on the basis of this concept they succeeded in achieving a fair degree of correlation with the scoring loads obtained at various speeds on their own gear testers.

However, in the first place their correlating method is of a highly empirical character, being but loosely bound up with the fundamentals of the viscoelastic behavior of oils. Moreover, with their correlation they aimed at explaining why scoring load went through a minimum at some characteristic speed of their test gears, so as to rise appreciably beyond that speed. But, judging from the theories hitherto available for accounting for the effect of viscoelastic behavior on the hydrodynamic load capacity of full fluid films, it would appear that this load capacity will not go up as speed increases but will rather go down. All in all, it would seem that a final conclusion about the worth of Borsoff's concept will have to await further developments in the application of viscoelastic theory to hydrodynamic lubrication.

A third possibility was suggested by Petrusевич (ref. 69), namely

---

<sup>44</sup> It would appear that, however high the maximum film pressure, oil will never solidify during its flow through a lubricant film, so that one need not here consider oils in a solidified state. In fact, times of exposure of oils to the high pressures in a film will be much shorter than the retardation times that are usual with nonadditive mineral oils subjected to solidifying conditions. But one may have to consider oils in a supercooled state, i.e., for as long as it takes the oil particles to flow through a film portion where the film pressures are higher than the solidification pressures at the film temperatures concerned.



that some critical shear stress, typical of the oil concerned, could never be exceeded because of some kind of fracture phenomenon that would occur even though the oil would remain in the liquid state. In the author's opinion, however, neither possibility lends itself to devising a mechanism that would satisfactorily explain the various recognized features of scoring.

A fourth possibility of breakdown of oils, which is common to all liquids that are not unusually free from certain nuclei, is that they cannot withstand any worthwhile subatmospheric pressures, let alone tensile stresses. In fact, before reaching zero absolute pressure the most volatile constituents of an oil, i.e., those having the highest vapor pressure, will evaporate so as to form bubbles in the oil film. This phenomenon will cause the oil to disperse into a state of cavitation. Such a state is indeed typical of the outlet portion of oil films between counterformal rubbing surfaces.

A fifth possibility which also results in a state of cavitation is afforded by the fact that oils in the state as supplied to the inlet of the film, even if devoid of aeration by bubbles of air, will usually be fully saturated with dissolved air, particularly when the lubricant supply system is of the circulating type. This involves that, as the oil temperature rises during the flow through the film, some fraction of the originally dissolved air is bound to be released in the film portion that precedes the outlet. The reason lies in the fact that the saturation limit for dissolved air will go up as pressure increases and will go down as temperature increases. Beyond the film cross-section where the oil pressure reaches its maximum, while film pressure falls and temperature still goes up, the oil must go again through its original saturation limit at some film cross-section where film pressure is still higher than the atmospheric one at the outlet cross-section. So, in the film portion between the two cross-sections concerned, release of some air, and thus cavitation, is bound to occur.

In practice, both possibilities for cavitation will usually occur simultaneously. Now, some kind of flow instability and collapse of the film is conceivable in that the disruption of the oil flow in the cavitation zone may be associated with a choking effect. Such an effect might influence the extent of the noncavitated zone and thereby the generation of film pressures and the resultant load capacity. But so far no satisfactory mechanism of scoring seems to have been established on the basis of cavitation or on an associated choking effect.

If, however, such a cavitation mechanism should be found to be actually significant for scoring, one might end up with a correlation between scoring temperatures of various nonadditive oils and some volatility criterion which might perhaps suitably be taken from the ASTM distillation curves of the oils concerned. Such a correlation would at least roughly be related to the one (see fig. 1) between scoring temperatures and vis-

cosity grades. After all, for mineral oils there is a roundabout correlation between viscosity grade and volatility.<sup>45</sup>

Finally it is worthy of note that the collapse of the oil film as caused by an adverse change in some oil characteristic will not necessarily result immediately in scoring. In fact, the incipient contact temperatures may be still lower than the scoring temperature. So, in general there will be a margin in the form of an intermediate regime of lubrication between the collapse and the onset of scoring.

#### Structural Instability of Full Fluid Films

Let us now turn to scoring mechanisms that belong to the second category where, without any significant change in the characteristics of the oil, the lubricant film breaks down as a structural element.

The earliest recognized and simplest kind of breakdown of a full fluid film as such is the one where the film becomes thinner and thinner as the operating conditions, such as load, become hydrodynamically less favorable. For instance, the load, as one of the hydrodynamic factors, may reach a critical limit beyond which the film becomes so thin that it can no longer provide complete separation of the two rubbing surfaces. Contact will then be established on certain asperities of the surfaces, but in its incipient and "grazing" stage it will be of a localized nature, being dispersed over the conjunction area.

True, the establishment of contact is one of the prerequisites for the occurrence of scoring. Incipient or "grazing" contact, however, will not invariably result in scoring, as will be obvious from the above considerations about the significance of incipient contact temperatures.

Thus the author does not agree with Borsoff et al. (refs. 70 to 72) who maintained that, at least in gears, scoring is invariably bound to occur upon incipient contact or, say, whenever elastohydrodynamic lubrication just becomes marginal. The author is inclined to leave this possibility open, but only in cases where sliding velocities are so high that the incipient contact temperatures at least reach the scoring temperature.

On the other hand, there is much evidence from both experiments and gear practice, for instance with low-speed wear (see fig. 8), that up to moderate sliding velocities contact may well occur without any kind of wear that might properly be called scoring. This implies that when load is increased either continuously or incrementally, particularly when sliding velocities are small, there will in general be a margin of increasingly severe partial elastohydrodynamic lubrication before scoring sets in.

---

<sup>45</sup> In this connection it is worthy of note that a few years ago the author's then collaborator, F. H. Theyse, found that on the Standard IAE 3¼ inch Centres Spur Gear Rig, the addition of a small amount of a comparatively volatile and hydrocarbon-type constituent to a nonadditive mineral oil resulted in a small but significant decrease in scoring temperature.



The existence of such a margin is, for instance, brought out by the experiments of Ibrahim and Cameron (ref. 73) insofar as these relate to small or moderate sliding velocities.

But these same experiments, like other evidence, are indicative of another possibility, namely that at sufficiently high sliding velocities the lubricant film may suddenly collapse while it is still comparatively thick. Such an event would be a most unpleasant surprise in that the expected separation of the two rubbing surfaces might seem to be quite safe from the point of view of incipient contact, let alone from that of scoring. It would appear that such a collapse has to be attributed to some kind of instability of the oil film as a structural element. Since nothing much is known with certainty about instabilities that might result in the collapse concerned, the author will tentatively put forward two hypothetical instabilities, both being of a thermal nature. Both instabilities exemplify what may be called "structurally thermal overstressing."

One of these instabilities, which is conceivably due to an interaction between the viscous shear stresses, the viscous heating, and the variation of the viscosity of the given oil with the variation of temperatures inside the film, will be treated in Appendix E.

The other instability, which is conceivably due to an interaction between the aforementioned thermal bulging effect on the rubbing surfaces and certain heating phenomena in the film interposed between these surfaces, will not be dealt with. The reason is that the theories needed have not yet been combined to a sufficient extent.

#### **Scoring Mechanisms Characteristic of Boundary and Natural Oxide Layers**

Even the pure hydrocarbon constituents of nonadditive mineral oils, the lubricants to which we shall again confine ourselves, are surface-active in the sense that they can be physically adsorbed<sup>46</sup> to the rubbing surfaces or, rather, to the thin oxide layers that naturally cover such surfaces in engineering practice. In addition to the pure hydrocarbon constituents, the mineral oils will, in general, contain small amounts of sulphur or oxygen compounds. When these compounds are absorbed they tend to be more firmly bonded, the reason being that they may react chemically in situ with the metals and oxides of the rubbing surfaces. But, it would appear that these naturally occurring chemically active constituents are much less effective than EP additives promoting the protection from scoring. In any case, in gear experiments or practice the author has never come across cases where differences in the scoring performance of

---

<sup>46</sup> In working out the theoretical equations for the full hydrodynamic lubrication, it is generally assumed that the flow in full fluid films is controlled, among other influential factors, by the absence of slip with respect to either rubbing surface. It would appear that this absence of slip may be attributed to the absorption of constituents of the oil to either surface.



nonadditive mineral oils could be ascribed systematically to differences in the amount and/or composition of the naturally occurring chemically active constituents.

It is well known that the natural oxide layers all by themselves, i.e., in dry sliding, provide a limited protection from scoring. This protection may be appreciably enhanced by the boundary layer that may be deposited from the adjacent and wetting oil.

As has already been set forth, the beneficial effect of both the oxide and the boundary layer on scoring may be sought in the contamination, and thereby the weakening, of the tiny welds that will form between the two virgin metals wherever these layers are punctured. Whenever the weakening is sufficiently effective, gross damage in the form of metal transfer, which above has been considered characteristic of scoring, may be prevented.

The effect of the weakening will be twofold. First, the inevitable tearing will take place only at the interface of the weld so that metal transfer is prevented. Second, the weld-shearing forces, and thus the sliding friction, will tend to be kept down to a reasonable level. Therefore, unless the sliding velocity attains a critically high value, the flash temperatures will also be kept from getting so high that the increase in the welding tendency of the two metals more than offsets the weakening effect through contamination.

In full fluid film-lubrication, both rubbing surfaces are provided with triple protection from scoring. To begin with, the full fluid film, all by itself and while it lasts as such, gives full protection from any contact whatsoever, and thus from scoring. But, both the boundary and the oxide layer still play a useful supplementary protective role, i.e., whenever the full fluid film should break down, through some such collapse mechanism as described above or in Appendices F and G.

The critical conjunction temperature, up to which the boundary layer and also the oxide layer may give adequate protection from scoring depends on the competition between two counteracting kinds of processes. These will be described later.

Shearing and tearing action during sliding tends to rupture boundary layers at least locally. The regenerative process or self-healing action is to be sought in the readsorption of surface-active constituents as soon as the oil gets access again to the damaged spots. But, it would appear that the self-healing action of boundary layers can no longer be very effective when the conjunction temperature exceeds the so-called desorption temperature, which is dependent on the nature of both the surface-active oil constituents and the rubbing metals, including the latter's oxides. In this connection it may be recalled that conjunction temperature consists of two components, the bulk temperature and the flash temperature component.

The scoring temperature, i.e., the critical conjunction temperature up to which protection from scoring persists, might conceivably vary with the relative contribution of the bulk temperature and the flash temperature component. This point may be illustrated by comparing the two following extreme cases.

Let the first case be characterized by operating conditions where the bulk temperature is equal to the desorption temperature, or very nearly so. This requires that the amount of sliding and thus the corresponding flash temperature component required to reach the desorption temperature, be minute.

At first blush this feature may look attractive for experimental determinations of scoring temperatures, at least when it is assumed that this temperature is identical with the desorption temperature. In fact, through utilizing small sliding velocities, an accurate determination is facilitated. Bulk temperatures can be readily assessable to a high accuracy while the demand for accuracy of the assessment of the comparatively small flash temperature component is practically eliminated. In the course of time, several investigators have followed up this line of experimental approach in slow speed four-ball machines and other simulating testers, including disk machines (Boerlage and Blok, ref. 74; Vinogradov, refs. 60 and 61; Matveevsky, ref. 65; O'Donoghue et al., ref. 75). But, in light of the second case now to be discussed, it remains to be seen whether the critical temperatures determined under such conditions and on such simulating testers may be considered really representative of the scoring temperatures in actual gear practice.

As a second illustrative case, let us now consider the case where the bulk temperature stays at levels that are distinctly lower than that of the desorption temperature. Accordingly a comparatively wide margin is left for frictional heating to provide a flash temperature component sufficing for reaching a critical temperature at which desorption and/or scoring will set in.

The present case, in contrast to the first, is characterized by the feature that only the fraction of the conjunction area where contact really occurs may be exposed to temperatures as high as the critical one, often only intermittently so. The surface temperatures must be equal to the bulk temperatures over the remaining fraction of the conjunction area so that the critical temperature is not reached. Moreover, the present fraction will be occupied by interstices in which oil is available at the comparatively low bulk temperature. Thus, there is a better opportunity, as compared with the first case, for comparatively cool oil to get access to spots where the boundary film has been desorbed and/or punctured, and thus to aid in self-healing of the boundary layer. Therefore, it might well be that in cases similar to the present one, the temperature at which scoring sets in, in the form of gross damage due to metal transfer, is not



identical with, but at least somewhat higher than, the desorption temperature as determined along the line of approach characterized by the first case.

Indeed, the author never was able to achieve, either with nonadditive mineral oils or with mineral oils containing polar additives, a workable correlation between his numerous scoring test results on spur and hypoid gears and the desorption temperatures on the Four-Ball Top (ref. 74). Further, he even found additives that on the Four-Ball Top lowered the desorption temperature but that in gears did not discernibly affect scoring performance, either adversely or beneficially. Moreover, tests on the same apparatus with various kinds of EP oils revealed that, just as with the aforementioned kinds of oils, temperatures could be assessed as critical in that friction rose markedly. These temperatures, however, could not be identified with scoring temperatures. In fact, they were markedly lower than even the lowest estimates that could be made of the scoring temperatures obtaining with the same EP oils in actual gears.

To say the least, the author is far from convinced that scoring temperatures representative of gears can be reliably assessed through the above-discussed method where they are put identical with the critical bulk temperatures determined on testers where sliding velocity, and thus the flash temperature component, is reduced to the utmost.

Let us now turn to other kinds of protective layers that may be deposited from the oil onto the natural oxide layers of the rubbing surfaces. It would appear that certain such layers may indeed aid in the protection from scoring.

For instance, Beane and Lawler (ref. 59) found on a gear tester on which the temperature of the oil in the circulation system could be adjusted up to levels appreciably higher than those in usual gear practice, that scoring loads with the nonadditive mineral oils investigated started decreasing when oil supply temperature was increased. But at roughly 200° C (400° F) a minimum was reached, and beyond that temperature the scoring load went up again.

It seems likely that at the higher supply temperatures certain constituents of these oils were oxygenated and that this resulted in the deposition of thin lacquer-like layers. Such layers may well have raised the conjunction temperature at incipient scoring beyond that at normal oil supply temperatures. Since under the rather severe oxygenating conditions concerned the composition of nonadditive mineral oils is bound to alter, it is then to be expected that the postulate about the constancy of scoring temperature will no longer be valid.

Some more light may well be shed upon the above question of oxygenation of oils by the enlightening experiments performed on a simulating tester by Vinogradov (refs. 60 and 61), who also pointed to some interesting effects of antioxidants in mineral oils.



All in all, it would appear that under the operating conditions normally encountered in gear practice (i.e., those where the author's postulate seems to be valid to a reasonable approximation), oxygenating effects are not so very important from the scoring point of view. So it is not thought that a scoring mechanism that might explain the postulate could be based on such oxygenating effects.

Another oxygenating effect is the one that may well be significant in the self-healing of the metal oxide layers that render some natural protection from scoring to rubbing surfaces. As has been pointed out by Lancaster (ref. 76), with such oxide layers, two counteracting processes may be considered.

One process is the local breakdown of an oxide layer by the shearing and tearing action due to sliding load. This process may be characterized by the rate of exposure of the virgin metals. The other process is the one of self-healing brought about by oxygenation of the virgin metals where these have been exposed, and may thus be characterized by the rate of oxygenation.

Conceiving both rates as temperature-dependent, it may be hypothesized that with an increase in conjunction temperature the rate of exposure increases more rapidly than that of oxygenation. So beyond a certain critical temperature, the former rate could no longer be offset by the latter. This critical temperature might then be conceived as the scoring temperature, in that, beyond it the self-healing would no longer suffice for preventing gross damage by metal transfer.

Lancaster's experiments, however, have been performed under dry-sliding conditions. Thus it remains to be seen to what extent one might uphold his conclusions when it comes to the significance of oxygenating effects for scoring temperatures under conditions usual for gear practice. In fact, in contrast to Lancaster's test conditions, the oxygen from the ambient air has no longer free access to the rubbing surfaces. It is admitted though that the ambient air still plays a part, albeit mainly in keeping the oil lubricating and covering the rubbing surfaces saturated with air or, say, with oxygen. Fair amounts of oxygen are dissolved in such saturated oils, but it would seem that the relative importance of self-healing by oxygenation under lubricated conditions differs much from that in Lancaster's dry-sliding experiments (see Vinogradov, refs. 60 and 61).

Again with nonadditive mineral oils, it seems likely that the postulate about the constancy of scoring temperatures can neither be explained on the basis of a competition between the self-healing and the breakdown, nor exposure, process of the natural oxide layers.

#### CONCLUSIONS AND RECOMMENDATIONS

As the author has pointed out previously, the postulate about the constancy of scoring temperature, expressed in terms of maximum surface

temperature in the conjunction zone of two rubbing counterformal surfaces, is not in general valid for those oils containing extreme-pressure additives.

With combinations of nonadditive mineral oils and rubbing counterformal surfaces made of steel, the comparatively great width of the postulate's applicability range is corroborated by published experimental evidence, provided this be interpreted in a sufficiently refined manner.

More refined interpretations, such as those illustrated in the various case studies of the third section, should preferably be extended both to future experimental work and that which is already available but which the author has not so far been able to analyze thoroughly. The additional knowledge so to be gained must come before limits to the postulate's applicability range can be more definitely set.

Among the various scoring criteria herein suggested, the author has not yet come across anyone showing a proven applicability range as wide as that of his own criterion—the scoring temperature as expressed in terms of maximum conjunction temperature.

Three major expedients for reducing the presently appreciable scatter in the correlation between scoring results on various gear testers, and also between these and those obtained on simulating testers such as disk machines, are

- (1) The introduction of, and adherence to, one single standard definition of scoring which should relate to the onset of this phenomenon and not to some post mortem stage

- (2) The introduction of improvements in both the control of, and knowledge about, the bulk temperatures in the various testers

- (3) Conversion of the conventional ratings of scoring performance, such as those in terms of scoring load, into scoring temperatures

In testers run under sufficiently controlled conditions, the scatter in scoring temperatures assessed through the most refined methods available would not appear to be greater than that normally considered acceptable for the results of fatigue tests representative of the pitting (surface fatigue) or the bulk fracture of gear steels.

Even though the postulate is not in general applicable to extreme-pressure oils, its wide applicability to nonadditive mineral oils is considered a real boon to gear designers. In fact, they are first in need of determining whether or not some nonadditive mineral oil will suffice for adequate protection from scoring.

There is no certainty yet about a scoring mechanism or a combination of such mechanisms that might satisfactorily explain the wide applicability of the present postulate. Therefore, several such mechanisms have been submitted for discussion.



**APPENDIX A. THEORETICAL DERIVATION OF THE THERMAL CONSTRICTION RESISTANCE OF A BAND-SHAPED CONJUNCTION ZONE**

Earlier the thermal constriction resistance of conjunction zones was discussed as one of the concepts basic to thermal networks representative of gear transmissions. A derivation will now be given of equations (11a) and (11b) there given for the thermal resistances,  $R_{c,1}$  and  $R_{c,2}$ , of the two rubbing bodies 1 and 2, respectively, that extend on either side of the conjunction zone.

The present derivation will be based on the flash temperature theory. This implies that the surface of either rubbing body in so far as it lies outside the conjunction zone, and particularly the portions of that surface that are adjacent to this zone, are assumed to be perfectly heat-insulated. Thus, the thermal constriction resistance concerned is characteristic of the feature that in flash heating the conjunction zone offers a constriction to the heat flowing from that zone, through the so-called "thermal skin," into the body concerned. By incorporating, into the thermal network representing a given gear transmission, the two thermal constriction resistances that are associated with a given conjunction zone, the effect of this zone on the bulk temperatures of all the various components of the transmission can be accounted for (see figs. 6 and 7).

Under conditions conducive to scoring in gears or disks, the instantaneous tangential velocities of the two rubbing surfaces with respect to their instantaneous conjunction zone are comparatively high. Practically speaking, it is then typical of flash treating that the portion of either rubbing body that lies instantaneously ahead of the leading edge of the conjunction zone still retains its original temperature. This temperature may thus be put equal to the bulk temperature of the body concerned. In other words, practically none of the heat penetrating from the conjunction zone will reach the portion concerned. But if the portion of the surface that lies instantaneously ahead of the conjunction zone is sufficiently cooled such as by an oil jet (ref. 18), one will have to account for diversion of part of the heat flow from the conjunction zone. Then, however, the surface temperature measured ahead of the leading edge of the conjunction zone will no longer be equal to the bulk temperature. The diverted portion of the heat will have to flow in a pattern somewhat similar to the one typical of the above-mentioned thermal skin. Hence, an estimate of the difference between the surface temperature concerned and the bulk temperature may be obtained, at least as to the order of magnitude, through utilizing a thermal "diversion resistance." The thermal resistance does not differ too much from the thermal constriction resistance now to be derived.

For deriving the thermal constriction resistance of a band-shaped conjunction zone, first introduce the concept of thermal conductance, which



is defined as the reciprocal of thermal resistance. Then, the total conductance,  $1/R_c$ , of the entire band-shaped moving conjunction zone, of width  $w$ , may be conceived to be built up of the elementary heat conductances,  $d(1/R_c)$ , of parallel strips of width  $dx$ , that together cover the entire width  $w$  (see fig. 9).

The conduction,  $d(1/R_c)$ , of an arbitrary strip  $dx$ , i.e., at an arbitrary distance  $x$  from the leading edge, 0, of the moving conjunction zone and of a length  $B$  which is equal to the width of the tooth face concerned, is readily found from the reciprocal of equation (10):

$$d\left(\frac{1}{R_c}\right) = \frac{1}{B} \cdot \frac{q \cdot dx}{T} \quad (16a)$$

where  $T$  is the local flash temperature and  $q$  the local heat flux, as defined for the semielliptical (Hertzian) distribution concerned, by:

$$q = q_{\max} \cdot 2 \left( \frac{x}{w} - \left( \frac{x}{w} \right)^2 \right)^{1/2} \quad (16b)$$

where  $q_{\max}$  is the maximum heat flux, which occurs at the center,  $x = \frac{1}{2}w$ , of the conjunction zone, and which is equal to:

$$q_{\max} = \frac{4}{\pi} \cdot \frac{Q}{w} \quad (16c)$$

where  $Q$  is the heat withdrawn per unit width,  $B$ , of the tooth face to be considered.<sup>47</sup>

Since the conductances of the elementary strips are in shunt, their resultant, the conductance  $1/R_c$  sought, may simply be found through a summation procedure, i.e., by integrating equation (16a) as follows:

$$\frac{1}{R_c} = \frac{1}{B} \int_0^w \frac{q \cdot dx}{T} \quad (17)$$

For the cases to which the author will confine himself, i.e., where both Péclet numbers,  $V_1 w / a_1$  and  $V_2 w / a_2$ , are much greater than unity, it is known from the original flash temperature theory (ref. 1) that the local flash temperature at a distance  $x$  from the leading edge may be found from the following integration:

$$T = \frac{1}{\sqrt{\pi}} \cdot \frac{1}{b \sqrt{V}} \int_0^x \frac{q}{\sqrt{x-\xi}} \cdot d\xi \quad (18)$$

After substituting equation (16b), this integration yields, for the range concerned,  $0 \leq x \leq w$ ,

<sup>47</sup> At this stage the subscripts 1 and 2 are not yet introduced to distinguish between the meshing tooth faces, 1 and 2.

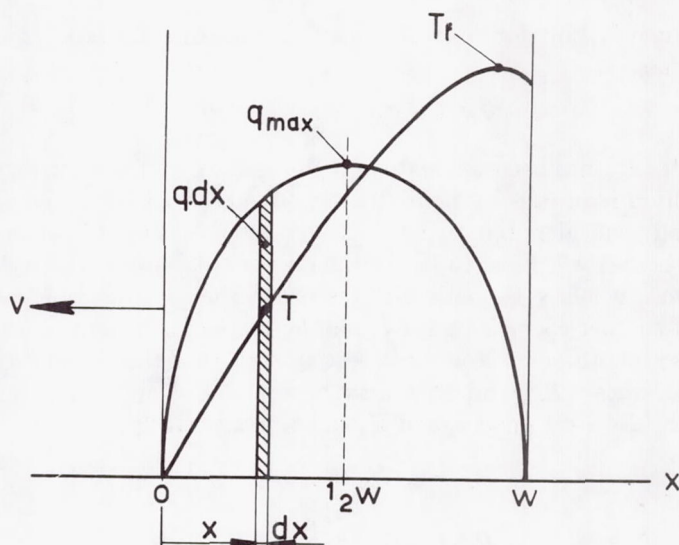


FIGURE 9.—Basic data for the thermal conductance,  $1/R_c$ , of a moving band-shaped conjunction zone of width  $w$ .

$$T = \frac{q_{\max} \cdot \sqrt{w}}{b \sqrt{V}} \cdot F\left(\frac{x}{w}\right) \quad (19a)$$

where

$$F\left(\frac{x}{w}\right) = \frac{4}{3\sqrt{\pi}} \left[ \left(1 - \frac{x}{w}\right) \cdot K\left(\sqrt{\frac{x}{w}}\right) - \left(1 - 2\frac{x}{w}\right) \cdot E\left(\sqrt{\frac{x}{w}}\right) \right] \quad (19b)$$

where  $K$  and  $E$  denote the complete elliptic integrals of the first and second kind, respectively, and as defined by:

$$K(k) = \int_0^1 (1-t^2)^{-1/2} \cdot (1-k^2t^2)^{-1/2} \cdot dt \quad (19c)$$

and

$$E(k) = \int_0^1 (1-t^2)^{-1/2} \cdot (1-k^2t^2)^{1/2} \cdot dt \quad (19d)$$

where, by definition,

$$k^2 = \frac{x}{w} \quad (19e)$$

It is thus found that the thermal constriction resistance amounts to:

$$R_c = \frac{1}{2} \cdot (B \cdot b \sqrt{Vw})^{-1} \cdot \left\{ \int_0^1 \left[ \frac{x}{w} - \left(\frac{x}{w}\right)^2 \right]^{1/2} \cdot \left[ F\left(\frac{x}{w}\right) \right]^{-1} \cdot d\left(\frac{x}{w}\right) \right\}^{-1} \quad (20)$$

By numerical integration<sup>48</sup> of the integral in equation (20), it is finally found that:

$$R_c = 0.478(B \cdot b \sqrt{wV})^{-1} \quad (21)$$

Result (21) has been evaluated for the case where the semielliptic heat distribution indicated by figure 9 is wholly absorbed by the single semi-finite body considered in that figure. In reality the case of two such contacting bodies will have to be considered, one having a velocity  $V_1$  and the other a velocity  $V_2$ , both with respect to their conjunction zone. The thermal contact coefficients are  $b_1$  and  $b_2$ , respectively, and may, in general, also be different. Now, it is readily seen that the thermal constriction resistances,  $R_{c,1}$  and  $R_{c,2}$ , may be found by replacing  $V$  and  $b$  in equation (21) by  $V_1$  and  $b_1$  and  $V_2$  and  $b_2$ , respectively, so that:

$$R_{c,1} = 0.478(B \cdot b_1 \sqrt{wV_2})^{-1} \quad (22a)$$

and

$$R_{c,2} = 0.478(B \cdot b_2 \sqrt{wV_1})^{-1} \quad (22b)$$

Results (22a) and (22b) are identical with equations (11a) and (11b), respectively, so that proof of these formulas has now been delivered.

Finally, by substituting the following relationship of Hertz for the width,  $w$ , of the conjunction zone,

$$w = \left( \frac{32}{\pi} \cdot \frac{WR}{E_r} \right)^{1/2} \quad (23)$$

equations (22a) and (22b) may be converted to:

$$R_{c,1} = 0.268 \frac{E_r^{1/4}}{B} (R^{1/4} W^{1/4} \cdot b_1 \sqrt{V_1})^{-1} \quad (24a)$$

and

$$R_{c,2} = 0.268 \frac{E_r^{1/4}}{B} (R^{1/4} W^{1/4} \cdot b_2 \sqrt{V_2})^{-1} \quad (24b)$$

which are identical to equations (12a) and (12b), respectively.

#### APPENDIX B. CALCULATION OF THE MAXIMUM FLASH TEMPERATURE AS CORRECTED FOR BAND-SHAPED CONJUNCTION ZONES TAPERED BY MALDISTRIBUTION OF THE LOAD

The original flash temperature equation (1b) and its variants (2a), (2b), and (2c) have been established for a band-shaped conjunction zone

<sup>48</sup> For this numerical evaluation, in so far as it is to be carried out in the vicinity of  $k = (x/w)^{1/2}$ , it proves useful to use the following series expansion of the intergrand,  $(\pi k)^{-1/2} \cdot (1 - \frac{1}{8}k - \frac{1}{8}k^2 - 83/1024k^3 \dots)$ .



of uniform width,  $w$ . This means that these formulas need be corrected when the load borne by the conjunction zone is maldistributed so that this zone becomes tapered in that its width will vary along its length (see fig. 2). Accordingly, the load,  $W$ , per unit length will also vary over the length, i.e., with the local width,  $w$ , of the zone concerned.

Since in actual practice the variation of the width is very small indeed as compared with the length of the conjunction zone, a relationship following from Hertz's theory for zones of uniform width, may here be applied as a relationship between the local width,  $w$ , and the local unit load,  $W$ , (compare with eq. (23) in Appendix A),

$$W = \frac{\pi}{32} \cdot \frac{E_r}{R} \cdot w^2 \quad (25)$$

Now, again on account of the comparatively slight variation of the width and, therefore, of unit load  $W$ , one may still apply the original flash temperature equations (1b) or (2a) for finding the maximum flash temperature,  $T_f$ , provided that the maximum unit load,  $W_{\max}$ , which occurs where width  $w$  is at its maximum,  $w_{\max}$ , is substituted for  $W$ , together with  $w_{\max}$  for  $w$ , in these formulas. Since in these formulas  $W$  and  $w$  appear in their combination  $W/w^{1/2}$ , the actual maximum flash temperature will be  $(w_{\max}/w)^{3/2}$  times the one calculated from the formulas concerned when, as is usual,  $W$  and  $w$  therein are taken as the uniform unit load and the uniform width of the conjunction zone for uniform distribution of the load. Thus proof of the assertion that immediately follows figure 2 has been delivered.

#### APPENDIX C. GENERALIZATION OF THE FLASH TEMPERATURE EQUATION FOR UNEQUAL BULK TEMPERATURES

The flash temperature theory, which originally was worked out for the particular, though not infrequent, case of equal bulk temperatures of the two rubbing bodies, has been given in a form generalized for unequal bulk temperatures,  $T_{b,1} \neq T_{b,2}$ . This has been done in the approximate form of equation (4b) for conjunction temperature,  $T_c$ . In the present appendix a more accurate formula will be derived and it will be shown that, in range (4a) for  $n = b_1 \sqrt{V_1}/b_2 \sqrt{V_2}$ , the comparatively simple equation (4b) indeed yields a good approximation.

In the present general case, the assumption about the absence of interfacial temperature jumps across the conjunction zone can be expressed by the following generalized version of equation (1a),

$$T_c = T_{b,1} + T_{f,1} = T_{b,2} + T_{f,2} \quad (26)$$

In accordance with the above assumption, this expression brings out that the conjunction temperature,  $T_c$ , is common to the two rubbing surfaces, 1 and 2. Further, whenever the two bulk temperature components,

$T_{b,1}$  and  $T_{b,2}$ , are unequal, the same is seen to be true of the two flash temperature components,  $T_{f,1}$  and  $T_{f,2}$ .

A particular case of unequal bulk temperatures has already been discussed. This case was particular in that both surfaces, or say bodies, were stationary with respect to their conjunction zone where they were held into contact from some initial instant onward. The solution for the so-called incipient contact temperature,  $T_i$ , has been given by equation (9a). Of course, starting at the initial instant, heat will flow through the conjunction zone and from the hotter to the cooler body. When from the outside no heat is supplied to, or withdrawn from, the system of the two bodies, the heat flow through the zone concerned will continue until these bodies have assumed the same temperature throughout.

A more complicated problem is the one where either body moves with respect to the conjunction zone and the two bulk temperatures, that is, the temperatures on either side of, and at a sufficiently great distance from, the conjunction zone are kept constant in some way or another, while no frictional heat is developed in that zone. Again heat will flow from the hotter toward the cooler body through the conjunction zone. It can be shown that this heat flow, unless the difference  $|T_{b,1} - T_{b,2}|$  of the two bulk temperatures is unusually high, tends to be very small as compared with the frictional heat flow,  $Q$ , that will be generated in conjunction zones operating under conditions representative of gears that are liable to scoring. Therefore, in what follows we shall neglect the former heat flow altogether. Accordingly the total heat flow through the conjunction zone is put equal to the frictional heat flow,  $Q$ .

Under the above assumption two heat flows,  $Q_1$  and  $Q_2$ , will flow from the conjunction zone in opposite directions, the former into body 1 and the latter into body 2. Their sum will be equal to the frictional heat,  $Q$ , as expressed by

$$Q_1 + Q_2 = Q \quad (27)$$

As in the original flash temperature theory, where the bulk temperatures had been taken equal, the main problem consists in determining the partitioning of the frictional heat,  $Q$ , among the two bodies, i.e., in determining  $Q_1$  and  $Q_2$ . This problem can again be solved on the assumption of the absence of interfacial temperature jumps across the conjunction zone. The only complication is that, because of the present inequality of the two bulk temperature components,  $T_{b,1}$  and  $T_{b,2}$ , the two flash temperature components,  $T_{f,1}$  and  $T_{f,2}$ , will also be unequal.

Now, in accordance with the original flash temperature theory,  $T_{f,1}$  and  $Q_1$ , and also  $T_{f,2}$  and  $Q_2$ , can be related as follows:

$$T_{f,1} = \frac{1.11}{\sqrt{w}} \cdot \frac{1}{b_1 \sqrt{V_1}} \cdot \frac{Q_1}{B} \quad (28a)$$

and

$$T_{f,2} = \frac{1.11}{\sqrt{w}} \cdot \frac{1}{b_2 \sqrt{V_2}} \cdot \frac{Q_2}{B} \quad (28b)$$

By combining equations (28a) and (28b) with equation (26), plain algebra yields the following expression for  $T_c$ , the maximum conjunction temperature sought:

$$T_c = 1.11 \frac{fW | V_1 - V_2 |}{(b_1 \sqrt{V_1} + b_2 \sqrt{V_2}) \sqrt{w}} + \frac{1}{2} (T_{b,1} + T_{b,2}) + \frac{1}{2} \cdot \frac{n-1}{n+1} \cdot (T_{b,1} - T_{b,2}) \quad (29a)$$

where  $n$  is defined by:

$$n = \frac{b_1 \sqrt{V_1}}{b_2 \sqrt{V_2}} \quad (29b)$$

It is readily shown that in the comparatively wide range (4a), that is,

$$\frac{1}{5} \leq n \leq 5 \quad (30a)$$

equation (29a) may be replaced by the following one to a good approximation:

$$T_c = 1.11 \frac{fW | V_1 - V_2 |}{(b_1 \sqrt{V_1} + b_2 \sqrt{V_2}) \sqrt{w}} + \frac{1}{2} (T_{b,1} + T_{b,2}) \quad (30b)$$

In fact, in range (30a) the last term in equation (29a) lies in the comparatively narrow range from  $-\frac{1}{3}(T_{b,1} - T_{b,2})$  to  $+\frac{1}{3}(T_{b,1} - T_{b,2})$ . Thus, it is seen that this term will usually be much smaller than, and accordingly may be neglected relative to, the term  $\frac{1}{2}(T_{b,1} + T_{b,2})$  which has been retained in the approximative equation (30b). Finally it need only be observed that equation (4b) is identical to the above-derived equation (30b).

#### APPENDIX D. DERIVATION OF THE APPROXIMATE EQUALITY OF $T_{i,\max}$ AND $T_f$

In figure 4 and the text belonging thereto, a treatment has been given of the problem of the incipient contact temperatures,  $T_i$ , arising on the occasional and intermittent contacts in marginal elastohydrodynamic lubrication. In that text the conclusion has been put forward that in that regime of lubrication these incipient or "grazing" contact temperatures can be calculated to a reasonable approximation through utilizing the original flash temperature equation (2a), therein substituting for  $fW | V_1 - V_2 |$  the heat actually developed per unit length in the fully elastohydrodynamic film. In other words, it would be permissible to put incipient contact temperature,  $T_{i,\max}$ , equal to  $T_f$  as following from that formula.



It is the purpose of the present appendix to prove the above conclusion. To this end the basic equation (9a) will be utilized, and mainly so in its specialized form of equation (9b) which corresponds with the specialization to equal thermal contact coefficients. The author will confine himself to this specialization since it was introduced by Manton et al. (ref. 42) in their work on the distributions of the surface temperatures,  $T_1$  and  $T_2$ , of the two rubbing surfaces operating under conditions of full elastohydrodynamic lubrication, i.e., in the absence of any contact whatsoever.

Either surface temperature distribution,  $T_1$  or  $T_2$ , can be expressed in terms of the distribution of the heat that is withdrawn per unit length and per unit time by the rubbing surface, 1 or 2, respectively, as follows:<sup>49</sup>

$$T_1 = C_1 \left( \frac{x}{w} \right) \cdot \frac{w^{1/2}}{b_1 \sqrt{V_1}} \cdot \frac{Q_1}{w} \quad (31)$$

and

$$T_2 = C_2 \left( \frac{x}{w} \right) \cdot \frac{w^{1/2}}{b_2 \sqrt{V_2}} \cdot \frac{Q_2}{w} \quad (32)$$

where  $C_1(x/w)$  and  $C_2(x/w)$  are functions of the dimensionless coordinate,  $x/w$ , in the direction of the tangential velocities,  $V_1$  and  $V_2$ , thus describing the shape of the surface temperature distributions,  $T_1$  and  $T_2$ . Further,  $C_1(x/w)$  and  $C_2(x/w)$  depend solely on the shapes of the distributions of the heats,  $Q_1$  and  $Q_2$ , withdrawn by the rubbing bodies 1 and 2, respectively, per unit time and per unit length of the conjunction zone or, say, per unit width of the meshing tooth faces or disks. It is noteworthy that  $Q_1/w$  and  $Q_2/w$  represent the average amounts of heat withdrawn per unit time and per unit area of the conjunction zone.

Since it is typical of elastohydrodynamic films that the heat carried away from the oil by convection and longitudinal heat conduction in the film itself is negligible, it follows that one may put to a good approximation:

$$Q_1 + Q_2 = Q \quad (33)$$

where  $Q$  denotes the total heat developed in the film per unit time and per unit width,  $w$ , of its conjunction zone.

Introducing the partition ratio:

<sup>49</sup> Equations (31) and (32) follow, just as the original flash temperature equation (2a), from the flash temperature theory for the cases to which the author will again confine himself, i.e., to those where both Péclet numbers,  $V_1 w/a_1$  and  $V_2 w/a_2$  are sufficiently high,  $w$  denoting the width of the conjunction zone and  $a_1$  and  $a_2$  the thermal diffusivities of the two rubbing materials. In this connection it is noteworthy that the flash temperature theory yields relationships, like equations (31) and (32), between the distribution of the heat and the corresponding distribution of surface temperature, that are valid irrespective of whether the heat originates from boundary friction or from viscous shear in the adjacent hydrodynamic film (see also footnote 13).

$$\frac{Q_2}{Q_1} = \gamma \quad (34)$$

as utilized by Manton et al. (ref. 42), one may write, in accordance with the heat balance as expressed by equation (33),

$$\frac{Q_1}{Q} = \frac{\gamma}{1+\gamma} \quad (35a)$$

and

$$\frac{Q_2}{Q} = \frac{1}{1+\gamma} \quad (35b)$$

Substituting equations (35a) and (35b) into equations (31) and (32), and specializing these formulas for the particular cases considered by Manton et al., i.e., where the thermal contact coefficients,  $b_1$  and  $b_2$ , of the two rubbing materials have the same value, it is found that:

$$T_1 = C_1\left(\frac{x}{w}\right) \cdot \frac{1}{b\sqrt{V_1}} \cdot \frac{\gamma}{1+\gamma} \cdot \frac{Q}{\sqrt{w}} \quad (36a)$$

and

$$T_2 = C_2\left(\frac{x}{w}\right) \cdot \frac{1}{b\sqrt{V_2}} \cdot \frac{1}{1+\gamma} \cdot \frac{Q}{\sqrt{w}} \quad (36b)$$

Let us now prove, at least for the particular cases where  $b_1 = b_2 = b$ , that the initial contact temperature,  $T_i$ , as calculated from the original flash temperature formula, may indeed be put equal to a reasonable approximation to the arithmetic average of  $T_1$  and  $T_2$ , as has been expressed by equation (9b) which is here represented in its alternative form:

$$\frac{\frac{1}{2}(T_1 + T_2)}{T_i} \cong 1 \quad (37)$$

Now, according to the flash temperature theory and substituting  $T_i$  for the flash temperature at the point  $x/w$ , it is found that:

$$T_i = C\left(\frac{x}{w}\right) \cdot \frac{1}{b(\sqrt{V_1} + \sqrt{V_2})} \cdot \frac{Q}{\sqrt{w}} \quad (38)$$

where  $C(x/w)$  is a function not only of the dimensionless coordinate,  $x/w$ , but also of the shape of the distribution of the total heat,  $Q = Q_1 + Q_2$  (see eq. (33)).

In general the shapes of the distributions of  $Q_1$ ,  $Q_2$  and  $Q = Q_1 + Q_2$  are different. Accordingly, the functions  $C_1(x/w)$ ,  $C_2(x/w)$ , and  $C(x/w)$  or, say, the shapes of the distributions of the surface temperatures,  $T_1$ ,  $T_2$ ,

and  $T_i$ , will be different. But, judging from figures 3, 4, 7, 8, and 9 of Manton et al. (ref. 42), the shapes of the distributions of the elastohydrodynamic surface temperatures,  $T_1$  and  $T_2$ , are very similar in each individual case covered in their comparatively wide operating range. Conversely it may be concluded from equations (31) and (32) that the following equality must have held good to a reasonable approximation throughout that range:

$$C_1 \left( \frac{x}{w} \right) \cong C_2 \left( \frac{x}{w} \right) \quad (39a)$$

or, say, that the distributions of the heats  $Q_1$  and  $Q_2$  must have been reasonably similar in any individual case of Manton et al. (ref. 42). It follows that the distribution of the total heat,  $Q = Q_1 + Q_2$ , must have been similar to that of either  $Q_1$  or  $Q_2$ , so that one may put more generally:

$$C_1 \left( \frac{x}{w} \right) \cong C_2 \left( \frac{x}{w} \right) \cong C \left( \frac{x}{w} \right) \quad (39b)$$

Substituting the equalities (39b) into equations (36a), (36b), and (38), it is a matter of plain algebra to derive the following formula:

$$\frac{\frac{1}{2}(T_1 + T_2)}{T_i} \cong 1 - \frac{1}{2}[1 - (1-s)^{1/2}] \cdot \left\{ 1 - \frac{2}{1+\gamma} \cdot \frac{1}{(1-s)^{1/2}} + \frac{1}{1+\gamma} \cdot \frac{1}{(1-s)^{1/2}} \cdot [1 - (1-s)^{1/2}] \right\} \quad (40a)$$

where  $\gamma$  represents the partition ratio, as defined by equation (34), and  $s$  the slide ratio, defined as follows by Manton et al.:

$$s = 1 - \frac{V_2}{V_1} \quad (40b)$$

Now, in the range covered by Manton et al. (compare their fig. 12), i.e.,

$$0 \leq s \leq 0.3 \quad \text{and} \quad 1 \leq \gamma \leq 11 \quad (41)$$

it proves that the second, rather complicated term on the righthand side of equation (40a) may be conceived as a comparatively small correction to the first term, unity. In fact, throughout range (41) this correction term does not exceed 0.067, or 6.7 percent.

It has thus been demonstrated that, throughout Manton's range (41), which is fairly representative of usual gear practice, the distribution of the incipient contact temperatures,  $T_i$ , like that of the very similar distribution of the arithmetic average,  $\frac{1}{2}(T_1 + T_2)$ , of the two elastohydrodynamic temperatures  $T_1$  and  $T_2$ , is fairly insensitive toward both the value of the sliding ratio,  $s = 1 - V_2/V_1$ , and the partitioning of the



total heat between the two rubbing bodies, as expressed by the partition ratio,  $\gamma = Q_2/Q_1$ .

However, the shape of these temperature distributions does depend on the shape of the total frictional heat,  $Q = Q_1 + Q_2$ , developed in the elastohydrodynamic film. Since the former shape is unknown a priori, so is the latter. But fortunately the particular incipient contact temperature that is most critical from the viewpoint of the risk of scoring, i.e., the maximum,  $T_{c, \max}$ , of the distribution of the incipient contact temperatures, can be shown as follows to be fairly insensitive toward the shape of the distribution of the total frictional heat,  $Q = Q_1 + Q_2$ .

The author has previously explored the maximum flash temperatures that are generated with various shapes of the distributions of the total frictional heat. He found that, in a wide range of such shapes, which also covers the shapes following from the work of Manton et al., the maximum incipient contact temperatures are rather insensitive toward these shapes and, for the given values of  $b$ ,  $w$ ,  $V_1$  and  $V_2$ , are dependent on the total frictional heat,  $Q$ . It follows that within the present range, the maximum incipient contact temperature,  $T_{i, \max}$ , may be put equal to a fair approximation to the maximum flash temperature,  $T_f$ , that will be reached for a Hertzian semielliptic distribution of the total frictional heat,  $Q$ .

Therefore, a fair estimate of this temperature may be made by utilizing the flash temperature equation (2a)<sup>50</sup> in the very simple manner of substituting the known value of the total frictional heat,  $Q$ , for the expression,  $fW |V_1 - V_2|$ , which appears in that formula. It does not matter that in many cases the total heat  $Q$  is not known at all, or at least not sufficiently reliably, from the thermal theory of elastohydrodynamic lubrication. In fact, in such cases one may still utilize equation (2a), provided that the coefficient of friction,  $f$ , is known experimentally. Since one may then substitute the known values of  $f$ ,  $W$ , and  $|V_1 - V_2|$  directly into equation (2a), calculation of  $Q$  from  $fW |V_1 - V_2|$  may even be dispensed with.

#### APPENDIX E. CONCEPT OF THERMAL COLLAPSE OF FULL FLUID FILMS BY VISCOUS HEATING

One of the two instabilities of a thermal nature that have been put forward earlier and that may hypothetically arise in full fluid films

<sup>50</sup> For a Hertzian semielliptic distribution of the total frictional heat the maximum flash temperature,  $T_f$ , will occur at about  $x \cong 0.825 w$ , putting the origin of the  $x$ -coordinate at the inlet edge of the conjunction zone. With distributions having different shapes the location of the maximum flash temperature will also be different, but its value,  $T_f$ , is nevertheless fairly insensitive toward these shapes as has been noted above. Thus, it follows from a comparison of the original flash temperature equation (2a) and equation (38) that, irrespective of its location the maximum flash temperature or, say, the maximum incipient contact temperature,  $T_{i, \max}$ , may also be found from equation (38) by substituting for  $C(x/w)$  the numerical factor 1.11 of equation (2a).

and make these collapse, is the one that might be related to the viscous heating that will take place throughout any such film. Although a complete theory on this particular kind of instability is not yet available, it is deemed worthwhile to discuss the underlying concept.

In laminar flow the heat,  $Q_v$ , developed per unit volume and per unit time by the local viscous shear may be put equal to

$$Q_v = \eta \left( \frac{du}{dz} \right)^2 \quad (42a)$$

where  $\eta$  is the dynamic viscosity at the local temperature and pressure, and  $du/dz$  is the local rate of shear,  $u$  denoting the flow velocity in the main direction of flow and  $z$  a coordinate perpendicular thereto, i.e., in the same direction in which the thickness of the film is measured. Using Newton's definition of dynamic viscosity,

$$\eta = \tau \left( \frac{du}{dz} \right)^{-1} \quad (42b)$$

where  $\tau$  denotes the local shear stress, equation (42a) may be re-written as follows:

$$Q_v = \frac{\tau^2}{\eta} \quad (42c)$$

Accounting for the viscous heating, and sometimes also for the heating and cooling effected by compression and decompression, respectively, of the oil in the film (see footnote 19), various investigators theoretically evaluated temperature fields in elastohydrodynamic films (Crook, ref. 39; Cheng and Orcutt, ref. 77; Dowson and Whitaker, ref. 40; Manton et al., ref. 42). Of course, they accounted also for the dissipation of the heat, which takes place mainly by heat conduction across the film in that the major part of the heat developed in the film is withdrawn by the two rubbing surfaces that bound the film. Further, all of these investigators confined themselves to steady states where not only the flow of the oil in the film, but also the film profile, and the temperature field concerned are steady.

Now, for considering the hypothetical possibility of the thermal instability under discussion, one should investigate the effect of transient perturbations that may be superimposed upon such steady states. This has apparently not so far been done. But a few clues may be pointed out in the following.

For steady states, the aforementioned investigators found invariably that the maximum oil temperature did not occur on one of the rubbing surfaces; but at some point inside the film. Now, at that same point the film pressure will be known also from the solution concerned. Thus the local viscosity,  $\eta$ , at this point is also determined.



Further, the point concerned lies invariably in the conjunction zone of the film. In this zone, the shear stress may be shown to vary only inappreciably across the film. It may thus be assumed constant over any given cross-section of the film, i.e., in the above-defined  $Z$ -direction. In any case, the shear stress,  $\tau$ , at the point under consideration will have a certain value in the steady state concerned. This involves that the heat,  $Q_v$ , developed per unit volume by viscous shear at that same point, will also be known (see eq. (42c)).

Now, perturbations are conceivable where the shear stress,  $\tau$ , increases, if only temporarily, and where the film pressure at the cross-section concerned remains constant, or very nearly so. Then, the local viscosity can vary only in accordance with the temperature variation that might be due to the increase in  $\tau$ .

If it is assumed, as a start, that the dissipation of the heat from the point considered is not altered by the perturbation, it is readily seen that an increase in shear stress will entail an increase in temperature, i.e., a decrease in the local viscosity. However, an increase in shear stress,  $\tau$ , that goes hand in hand with a decrease in the local viscosity,  $\eta$ , will cause the heat generated locally,  $Q_v$ , to increase (see eq. (42c)). Such a chain of events may be conducive of a thermal instability, and thus of a thermal collapse of the film.

The crux of the present problem of thermal instability lies in the fact that the above assumption cannot in general be upheld. In fact, a perturbation in the shear stress may be expected to be associated with an equally transient perturbation in the local dissipation of the heat. Therefore, a refined analysis will have to be set up in order to find out definitely whether or not (and if so in what range of conditions), a thermal instability may indeed develop in that the tendency for local temperature to rise with an increase in the viscous heat  $Q_v$  is not offset by a simultaneous increase in the dissipation of the heat.

Unfortunately the refined analysis, on the basis of which the effect of perturbations could be studied, is even more complicated than the analyses that have been applied to steady states. However, it can be said that, owing to the transient character of the problem concerned, not only the heat conductivity of the oil but also its specific heat<sup>51</sup> will play a role in any criterion that might be found for the kind of thermal instability under discussion.

Because of the withdrawal of the viscous heat by the rubbing surfaces, the thermal contact coefficient of the rubbing materials may also be

---

<sup>51</sup> It is characteristic of steady states in representative cases of elastohydrodynamic lubrication that the heat dissipation by convection, where the flowing oil serves as a carrier of heat, is small as compared with the heat dissipation by the conduction of the heat across the film. Therefore, in contrast to the transient states the role of the specific heat may in steady states often be neglected.



expected to appear in such a criterion. Thus, characteristics of both the oil and the rubbing materials will be significant for the present thermal instability.

Finally it may be observed that if such a thermal instability should prove possible, it would result in a collapse of the film that might be termed a "thermal implosion of the film." Such an implosion would be the inverse of what happens in so-called "exploding wires" or "exploding foils," although all these cases would have in common that the instability is caused by the transient increase in heat generation that is not more than offset by the transient dissipation of the heat.<sup>52</sup>

#### DISCUSSIONS

**P. M. Ku** (Southwest Research Institute, San Antonio, Texas)

In his long and distinguished career, Professor Blok has made many great contributions in the fields of lubrication and machine design. His "flash temperature theory" is an excellent example of such contributions in that it has both important theoretical implications and powerful design consequences. In inviting Professor Blok to present a lecture on the flash temperature theory, the Steering Committee sought not only to honor a leading authority, but also to generate an interdisciplinary discussion on the theoretical and practical ramifications of a theory which has challenged the imagination of lubrication scientists and design engineers for three decades. Unfortunately, due to unforeseen but unavoidable circumstances, Professor Blok could not be present to deliver the lecture in person. The privilege of substituting for him thus became mine.

In my presentation of the flash temperature theory, I attempted to summarize, as best I could, the salient points so thoroughly covered in Professor Blok's text. But in line with the spirit of the symposium, I also took the liberty of posing some broad questions which I felt might help facilitate a better understanding and appreciation of the theory. This written contribution will not repeat what Professor Blok has already said. Rather, it will set down these broad questions, with the hope of soliciting Professor Blok's comments.

As Professor Blok has noted in the paper, he proposed the flash temperature theory 30 years ago. The theory has since found wide application in the design of lubricated gears and other types of counterformal contacts. It must be regarded as having stood the test of time, for even though its theoretical validity has been questioned, nevertheless, no equally powerful concept which takes into account the effects of lubricants and metals and other relevant variables has yet been advanced.

---

<sup>52</sup> The phenomenon of an exploding wire or exploding foil may be realized by releasing a sufficiently strong electric current through a metal wire or foil.

Now, quite apart from the design of gears or counterformal contacts, Professor Blok's analysis of the surface temperature profile in sliding systems from which the flash temperature theory evolved has, in my mind, great historic importance regardless of the significance of the flash temperature theory itself. Indeed, this pioneering work may justly be said to have laid the theoretical and mathematical foundation for the numerous sliding surface temperature investigations that have emerged in the past 30 years.

As I understand it, the flash temperature theory may be regarded as being composed of two parts: a general hypothesis and a specific corollary. The general hypothesis is that, in a sliding system under boundary lubrication conditions, scoring will occur when the maximum surface temperature in the conjunction reaches a critical magnitude. The specific corollary is that, for a nonadditive mineral oil and conventional gear steel, this critical temperature, or scoring temperature, is constant regardless of the operating conditions. Assuming that my interpretation is correct, then assessment of the validity of the theory rests basically on explaining what the scoring temperature is, rather than in establishing whether it is or is not constant under a restricted set of conditions. Of course, if the scoring temperature is constant for a nonadditive oil and conventional gear steel, the design analysis will be relatively simple. On the other hand, if the scoring temperature is not constant but its variation with any of the variables is known, the flash temperature theory will still be applicable even though the analysis will be more involved. Furthermore, the problem of great interest in modern industry is the behavior of EP-additive mineral oils and synthetic oils and unconventional steels, for which there is strong evidence that the scoring temperature is not constant (refs. 78 and 79). We need to learn more about such behavior. But even this does not necessarily invalidate the general hypothesis of the flash temperature theory, if only the scoring temperature can be given a rational explanation.

Let us now consider the general hypothesis. The existence of a critical failure temperature in a sliding system operating in the boundary lubrication regime, when friction rises abruptly and wear becomes significant, is an experimental fact. In carefully controlled experiments in which the sliding velocity is very low so that the temperature rise due to friction is negligible, this failure temperature can be related to the melting temperature of the lubricant-metal-atmosphere interaction product (refs. 80 and 81). It is this relationship that explains the failure mechanism and lends credence to the failure temperature under these conditions. But this failure temperature, as pointed out by Professor Blok, is associated with low-speed wear and not with scoring. The question then is what the scoring temperature is, or how can it be related to some fundamental phenomena. It appears to me that this is really



the key issue, and a critical examination of what is happening in the conjunction that leads to scoring is urgently needed.

Professor Blok has gone to great lengths to defend the specific corollary of constant scoring temperatures for nonadditive mineral oils and conventional steels. He has suggested numerous refinements to be made in experimental and analytical techniques; but he has not actually carried out such suggestions to the logical conclusion of proving the corollary. I do not mean by this statement that his ideas should not be followed up; but these are essentially peripheral issues. I cannot help but feel that the flash temperature theory does not rise or fall with the constant scoring temperature corollary for a nonadditive oil and conventional steel. Our rather extensive investigations in the past few years (which unfortunately we are not at liberty to publish at this time) show that even for nonadditive mineral oils and conventional gear steels, the scoring temperature is not constant with respect to some operating variables and surface roughness.

I believe the recent advances in elastohydrodynamics have generated information which underscores the need for an approach such as the flash temperature theory. With the means at hand, we can now define and predict with reasonable accuracy the conditions under which the operation moves from the elastohydrodynamic into the boundary lubrication regime. Thus in a situation totally devoid of chemical lubricant-metal-atmosphere interaction, elastohydrodynamics combined with a knowledge of surface topography provides a rational approach to design. However, this approach tends to be too conservative in the real situation where the consequences of the lubricant-metal-atmosphere interaction must be reckoned with. In other words, scoring does not necessarily take place the moment the operation enters the boundary lubrication regime, and how far the operation must extend into the boundary lubrication regime before scoring occurs depends upon the nature of the lubricant-metal-atmosphere interaction. In my way of thinking, it is because of this that some such approach as the flash temperature concept may give a more realistic design. Our own experience in this regard is that, with nonadditive mineral oils and common gear steels, the chemical effect is small and the elastohydrodynamic approach gives results that are no worse than the assumption of a constant scoring temperature. It is only when the chemical effect is large that the elastohydrodynamic approach breaks down. This being the case, the attractiveness of the flash temperature theory resides principally in situations where the chemical lubricant-metal-atmosphere interaction is significant and not in situations where such interaction can be neglected.

Professor Blok has emphasized, among other things, the need for a rational analysis of heat transfer in lubrication system design, and has suggested a thermal network approach. The problem of heat transfer



has scarcely received any attention in lubrication system design and in most of lubrication research. Professor Blok is to be commended for calling attention to an important and neglected problem. As a friend and admirer of Professor Blok, I cannot but marvel at his foresight and imagination.

**A. Dyson (Shell Research Limited, Thornton Research Center, Chester, England)**

I wish to make three comments on Professor Blok's paper. The first concerns the "thermal bulge" effect of Korovchinski. This analysis is valid only for the low velocity situation, when

$$\frac{VB}{a} \ll 1$$

where the notation of the paper has been used. The theory of Korovchinski is therefore inapplicable in the analysis of the contact conditions in most cases of practical interest.

The second comment concerns the effect of speed on the total contact temperature at scoring.\* A critical experiment was performed by Genkin, Kuz'min and Misharin (ref. 82), who ran a loaded pair of rolling and sliding discs until temperature equilibrium had been attained. They then reduced the rotational speed, keeping the load and the slide/roll ratio constant. As the speed decreased, the coefficient of friction increased, but the product of the sliding speed and the frictional traction decreased. Both the measured bulk temperature of the discs and the calculated flash temperature in the conjunction also decreased, and therefore the total contact temperature decreased continuously as the speed dropped. Nevertheless, under certain conditions, the discs scored. This demonstrates that the critical total contact temperature for scoring decreases with decreasing speed at constant slide/roll ratio, under the conditions of the experiment.

A similar conclusion may be drawn, rather more indirectly, from the many results available from scoring tests, both with discs (refs. 82 to 84) and with gears (refs. 15 and 85 to 87), which show the existence of a region in which the scoring load increases with increasing speed at constant slide/roll ratio. Again the coefficient of friction decreases with increasing speed and load, but it is an experimental observation that the product of the frictional traction and the sliding speed, i.e., the power dissipated in friction, always increases with increasing speed at constant slide/roll ratio. Both the bulk (blank) temperature and the flash temperature at scoring failure increase with increasing speed; therefore, the critical total contact temperature for scoring failure must increase with increasing speed at constant slide/roll ratio.

---

\* The term "scoring" is used in the American sense. The corresponding English term would be "scuffing."

The third comment concerns the relation between total contact temperature at scoring and the estimated thickness of the elastohydrodynamic film. Leach and Kelley (ref. 22) plotted these two quantities and concluded that the film thickness had no effect on scoring failure. Their work suggests inverse correlation between the two quantities. The total contact temperature just before scoring failure apparently decreases with increasing film thickness. This correlation becomes stronger if account is taken of the fact that the figures given for the highest film thicknesses are almost certainly over-estimates, whereas the estimates for the thinner films are probably accurate.

A similar correlation may be obtained from the results of Genkin, Kuz'min, and Misharin (ref. 82). Figure 10 shows a plot of the figures calculated from their experimental results for oil MC20 and steel 12X2H4A. The circled points denote a slide/roll ratio referred to the slower disc,  $U_1 - U_2/U_2$ , of 1.5, whereas the ratio for the uncircled points was 2.63. There is an undoubted correlation between total contact temperature and film thickness, but this becomes clear only if the points are separated according to the sliding speed. If Leach and Kelley's results were treated in the same way, the correlation might become more evident.

If this correlation has any physical significance, it would be expected that the critical total contact temperature should increase with increasing film thickness, and it therefore seems that no physical significance should be attached to the apparent correlation.

**A. Beerbower and F. F. Tao (Esso Research and Engineering Company, Linden, New Jersey)**

Professor Blok offers a number of provocative hypotheses to explain the "constancy of scoring temperature" shown in his figure 1. Actually this would be better described as "constancy of scoring temperature for a given lubricant" or "constant relationship of scoring temperature to kinematic viscosity at 60° C." The relationship takes the form of two lines,  $T_L = 98 \log K + 38$  and  $T_u = 121.5 \log K + 46$ , where  $T_L$  is the temperature (°C) for a very low probability of scoring,  $T_u$  the temperature for a very high probability of scoring, and  $K$  the viscosity at 60° C in centistokes. Since the figure also implies that the oils involved are all of 100 VI, it is possible to use chemical engineering correlations to test several of these hypotheses. The calculated properties are shown in table 1.

The procedure used was to select certain of Blok's oils, covering the range of data, and to use these as examples rather than manipulating the  $T_L$  and  $T_u$  equations. However, smoothed temperature values from the lines were used. The first step was to convert the viscosities and VI to the US customary temperatures of 100° and 210° F, using ASTM Method D2270 and Chart D341F. The latter was used to extrapolate

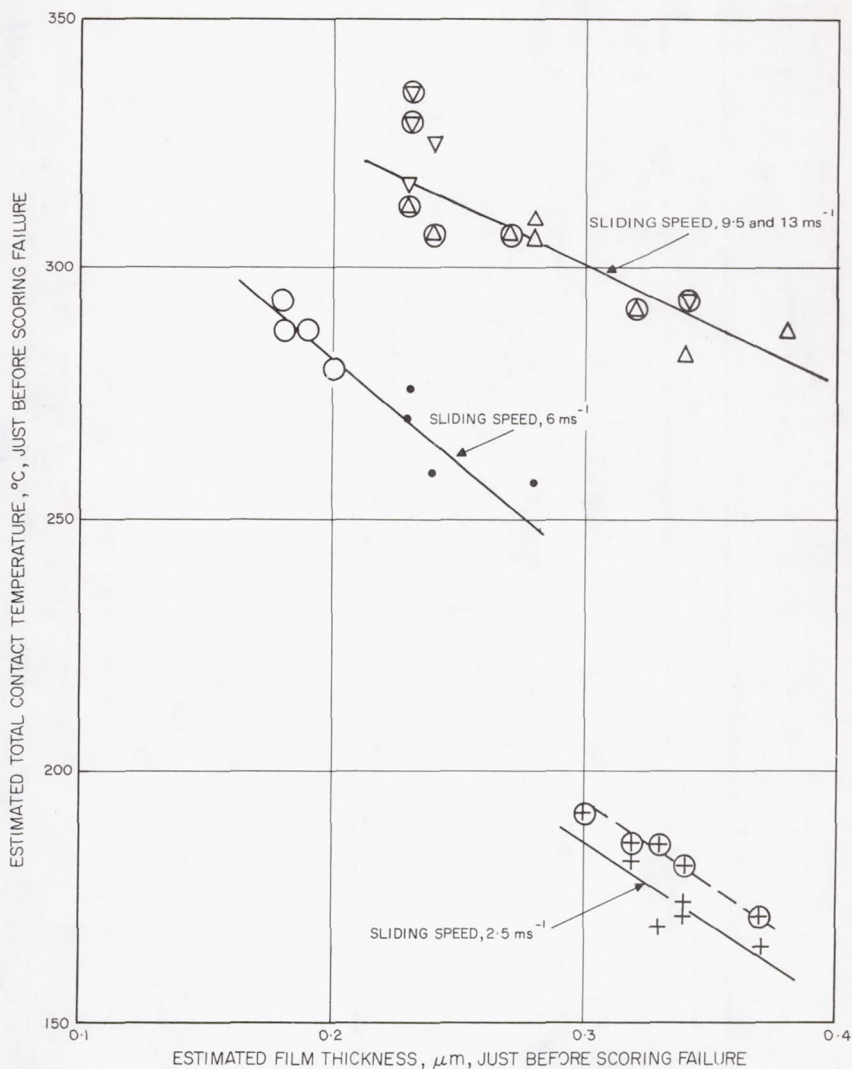


FIGURE 10.—Relation between total contact temperature and film thickness at scoring, from results of Genkin, Kuz'min, and Misharin (ref. 82).

these kinematic viscosities to the scoring temperatures. Densities at 60° F ( $d$ ) were obtained by ASTM Method D2501 for Viscosity Gravity Constant (VGC) and corrected to scoring temperature by API Technical Data Book Chart 6A3.4. This involved making the usually correct assumption that  $100 \text{ VI} = 0.800 \text{ VGC} = 12.00K$ , where  $K$  is the Watson Characterization Factor  $(T_b)^{1/3}/d$ . These corrected densities were used to convert the scoring viscosities to centipoises. The  $d$  values



TABLE 1.—*Estimated Properties of Blok's Oils at Standard Temperatures*

## A—Oil Properties

Oil no.	Viscosity, cs			Density at 60° F	Air solubility at 77° F and 760 torr, % by volume
	60° C	100° F	210° F		
1	7.2	14.1	3.2	0.843	12
2	14.5	33.3	5.3	0.857	11
4	31.0	82.8	9.5	0.872	9
5	82.0	249.7	20.0	0.884	8
7	145.0	509.0	32.0	0.889	8

B—At minimum scoring temperature ( $T_L$ )

	Temperature		Viscosity, cp	Vapor pressure, torr	Air solubility at 760 torr		Pressure of air plus vapor, torr	Surface free energy, ergs/cm <sup>2</sup>
	°C	°F			Vol %	Change from 77° F		
1	122	250	1.77	4	19	7	485	24.2
2	150	302	1.84	5	18	7	470	24.0
4	185	365	1.82	9	17	8	409	23.5
5	220	428	1.99	14	17	9	374	22.2
7	250	482	1.92	33	17	9	393	21.3

C—At maximum scoring temperatures ( $T_u$ )

1	150	302	1.24	10	20	8	470	22.6
2	190	374	1.16	30	21	10	430	21.9
4	230	446	1.15	52	19	10	412	20.8
5	280	536	1.17	105	18	10	445	19.6
7	309	588	1.17	175	18	10	515	18.5

were used with  $K$  to obtain the mid-boiling points ( $T_b$ ), from which 50° F was subtracted to approximate the 5-percent boiling point by ASTM D1160. These 5-percent points were used with the scoring temperatures to obtain the vapor pressures ( $p$ ), using API Technical Data Book Chart 5A1.15.

The surface free energies were calculated from an unpublished correlation developed under U.S. Army Contract DAHC19-69-C-0033, by means of pseudocritical temperatures obtained from API Technical Data Book Chart 4A1.2, and the  $T_B$  and  $p$  values. Air solubilities were estimated from ASTM D2779, extrapolated considerably beyond its scope but believed to be reasonably valid up to 700° F.

Professor Blok offers three categories of scoring mechanisms. Each of these has subcategories, some of which can and others cannot be analyzed by the properties estimated in table 1.

Professor Blok also lists five possibilities: (a) breakdown of viscosity at some critical shear stress or rate of shear, (b) development of viscoelasticity, (c) film fracture at some critical shear stress, (d) cavitation due to vaporization, and (e) cavitation due to air release. It is clear from table 1 that the most constant factor is the viscosity; the levels noted are those of kerosene which does not show any such anomalies as (a) or (c). The correlations shed no light on the viscoelastic theory, but it is extremely clear that the vaporization theory does not stand up to this test. Even allowing for some error of estimate, a "critical vapor pressure" which varies by a factor of 10 or more is not acceptable. The air solubility theory fares even worse, as this solubility actually increases at high temperatures. Furthermore, even adding the vapor pressure and the absolute pressure of the air (assuming saturation at 77° F and 760 torr) does not yield a constant, or distinguish consistently between  $T_L$  and  $T_u$ . It would appear that, were cavitation responsible for scoring, the critical condition might well be a minimum temperature!

In the main text and further in Appendix E, Blok treats certain possible cases of hydrodynamic instability. The basic case, loss of film thickness due to low viscosity, is usually attributed to decrease to a critical value of the Summerfield number ( $\eta V/W$ ), where  $V$  is velocity and  $W$  is load. This is inconsistent with the constant viscosities shown in table 1, as was shown by Fein (ref. 90). He also found a nearly constant viscosity for scoring transition, but in the four-ball machine on 52100 steel this was about 3 to 5 cps, just 2.7 times that shown in table 1. His attempt to correlate his data on the  $\eta V/W$  basis was not very satisfactory, as it shows an 800-fold span of this parameter to cover all the points.

Blok's reasoning in Appendix E is plausible but raises some questions. He states that "perturbations are conceivable where the shear stress

increases . . . and where the film pressure at the cross-section concerned remains constant, or nearly so." This seems promising, but vague. Does he mean that passage of a wear particle through the narrow gap might be enough to trigger off a tremendous increase in temperature?

Professor Blok raised questions about surface activity. The first two involve the concept of "desorption temperature," in which the lubricant abruptly loses its affinity for the gear surface. This concept is faulty, as the various forms of interfacial energies (work, heat, etc.) decrease continuously up to the vapor-liquid critical temperature. For the oils in question, this runs from 890° to 1050° F, far beyond the working range. It would be more logical to speak of a "displacement temperature" at which the work of adhesion of the atmosphere exceeds that of the liquid. However, the surface free energies of the oils shown in table 1 are far from low enough to permit such displacement. The apparent near-constancy of these values is deceptive, as it merely reflects the slow rate of change in surface properties with temperature as compared to the rapid rates of change of viscosity and vapor pressure. It should be mentioned that the calculation assumes that no help is to be had from "natural surfactants," and reflects only the surface energy of hydrocarbons.

Failure of a protective layer of oxygenated or other polymer has interesting possibilities but is beyond the reach of available correlations at present. Some light was shed on this by Fein (ref. 91), who found that break-in procedures strongly affect the scoring temperature and attributed this to varnish-like films detected after long break-in periods.

It might appear that Blok's final hypotheses on metal oxide films would also be outside the scope of this correlative analysis. However, if one considers the nearly constant amount of oxygen available in the solutions of table 1, it seems that the rate of self-healing of an oxide film would be essentially limited by oxygen diffusion (ref. 92) and hence dependent on both the viscosity of the oil and the temperature. This does not seem to lead to scoring at a fixed viscosity.

Thus none of Professor Blok's hypotheses fit his data without stretching them considerably. There is therefore a need for new hypotheses in order to explain the constancy of scoring viscosity. In view of the dependence of scoring temperature on viscosity, the new postulations may be sought from the standpoint of the major function of viscosity in lubrication, i.e., the load carrying effect. It is probable that among the asperity-contacts there are capillary channels through which the lubricant flows away as fast as the motion brings it in. The hydrodynamic pressure is developed due to the viscous drag of these capillary flows to counterbalance a part of the applied load. The net load on the asperities is thus the difference between the applied load ( $W$ ) and the integrated force of the hydrodynamic pressure ( $P$ ). According to the classical



hydrodynamic theory,  $P$  increases with the viscosity. This implies that the scoring temperature may be higher for a more viscous oil. This seems to involve a force equilibrium so that the viscosity at scoring is nearly a constant.

It should be emphasized that the narrow ranges of viscosity in table 1 are a tribute to Blok's refined method for prediction of flash temperatures.

**R. S. Fein (Texaco Research Center, Beacon, New York)**

Professor Blok indicates that workable correlations between the four-ball machine and spur gears have not been achieved. This seems to be a generally true statement. In fact, any correlation is sufficiently rare that it appears desirable to show that in at least one case a correlation can be obtained.

Figure 11 shows the correlation of critical temperatures obtained on a four-ball machine and gear scuff data obtained by Ibrahim and Cameron (ref. 73).<sup>\*</sup> Within the precision of the data,<sup>\*\*</sup> the four-ball and gear rigs give the same critical temperature dependence upon ratio of sum velocity to load.

It must be emphasized that this correlation probably resulted from the low speeds and blank temperatures of the I-C tests. More usually, as Professor Blok indicated, there is a lack of correlation that seems to result because usual gear test procedures produce extensive run-in during the conduct of the tests (ref. 89).

It seems clear from available evidence that so-called scuffing or scoring involves strong adhesion between the load-bearing surfaces and consequent metal transfer. It is also clear that strong adhesion requires close approach of atomically clean surfaces (ref. 93). Thus any scuffing mechanism must include a means for producing the atomically clean opposing surfaces and bringing them within close (i.e., atomic dimensions) approach.

Recent studies of adhesion and friction show the extreme difficulty of producing atomically clean surfaces even in the stringent environment of ultrahigh vacuum (ref. 94). Hence, in an oil atmosphere, production of sufficiently clean surfaces by "thermal desorption" appears unlikely. Rather, it seems necessary to physically disperse the contaminants separating the bearing surfaces by generating new surfaces by plastic deformation (ref. 93) as is commonly done in many welding processes.

---

<sup>\*</sup> The I-C data were kindly provided by Professor M. Ibrahim.

<sup>\*\*</sup> Four-ball critical temperatures were determined as indicated in reference 88. Gear critical temperatures were obtained from flash temperatures computed by Kelley's formula (ref. 23) and gear blank temperature assumed equal to the sump temperature; the latter assumption becomes less valid at higher sum velocities and provably accounts for increased data scatter.

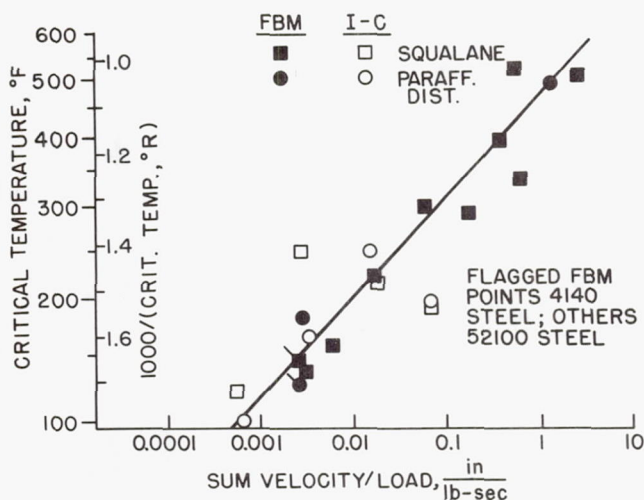


FIGURE 11.—Correlation of critical temperatures for four-ball and gear rigs.

The Tabor junction growth mechanism would appear to provide both for the plastic deformation and contaminant dispersal on bearing surfaces (ref. 95). Further, the likelihood of adhesion occurring as a result of junction growth should be sensitive to how much contaminant separates the surfaces before plastic deformation begins and to the extent of the plastic deformation. Viewed on this basis, one can at least rationalize why scuffing limits often seem to depend on a "critical" temperature (ref. 96). Thus the amount of lubricant separating asperities before plastic deformation must depend on the flow properties of the lubricant (whether liquid- or solid-like). These, in turn, are sensitive to temperature. Further, the onset of plastic deformation and its extent also will depend on temperature; thus temperature affects both the shear stress transmitted by the shearing lubricant film and the flow properties of the steel.

Therefore, it appears possible to develop a rational qualitative mechanism of scuffing that is consistent with modern knowledge of adhesion and the sometimes apparent existence of a critical temperature for scuffing.

**E. M. Kohn** (*Sun Oil Company, Marcus Hook, Pennsylvania*)

Professor Blok's concept of a critical temperature for scuffing implies that precipitous events surround the onset of scuffing. So far as the temperature is concerned, I would not expect it to rise sharply before scuffing. Until scuffing occurs, I would think that the temperature rises gradually and continuously. It may rise sharply to reach the critical



temperature after scuffing begins. But then, however, the critical temperature becomes an effect of scuffing rather than a cause of it.

Instead of characterizing scuffing in terms of a temperature, I think that scuffing might be tied to some temperature-related factor, or factors, which change sharply at some point in time. Factors to consider as the run progresses are (1) the drop in the hardness of the running surface and (2) the increase in the average contact area per weld which, of course, depends upon the surface hardness.

In a paper presented at this symposium (ref. 97), Zaretsky and Anderson show in figures 1 and 2 that the hot hardness of bearing steels drops rapidly over small temperature intervals. Let us apply this in the following mechanism. As the surface temperature rises progressively, the mechanical properties of the metal's surface deteriorate with time. (This picture is simplified by not including any side reactions such as the oxidation of iron to produce a surface hardened by the formation of any iron oxide.) The rise in temperature increases both the average area per junction and the volume associated with this area. As the junction volume increases, so does the heat produced by the shear of the metal in the volume. This feedback continues between an increasing junction volume and the increasing heat output per sheared junction until an average surface temperature is reached where the junction has attained a critical size. At this point, metal is plucked away and, as an aftermath, the heat produced during the plucking causes the surface to reach the critical temperature.

I think that pre-scuffing conditions should be evaluated experimentally in terms of the growth of contacts with running time (perhaps by the method of electrical contacts).

#### LECTURER'S CLOSURE

It is thought that Mr. Kohn's suggestions may well bear fruit in clearing up the mechanism(s) that result in scoring. In following up these and similar suggestions, however, one would have to distinguish carefully between, on the one hand, true scoring, in the form of the gross and self-aggravating damage as meant both in the aforementioned definition of the Institution of Mechanical Engineers and in the author's postulate, and on the other hand scoring of the self-healing type, not to forget types of wear having nothing to do with scoring (see fig. 8). The author, as is evidenced by his extensive survey of conceivable scoring mechanisms, agrees with Mr. Kohn that a clearer understanding of precipitous phenomena surrounding scoring might well prove very helpful. In this connection the author points to the need of rationally specifying the limits to the range of applicability of the postulate that must undoubtedly exist but that, to the best of his knowledge, have not so far been disclosed in a way workable to oil technologists, metallurgists,



and/or machine designers. Finally it is worthy of note that in the disc tests of Leach and Kelley (ref. 22), the method of electrical contacts has already proved its worth for a precise assessment of the onset of scoring.

Dr. Fein suggests a scoring mechanism that is characterized by the essential role assigned in adhesion, or say scoring, to junction growth as studied by Tabor (ref. 95). It is thought that Dr. Fein's suggestion is certainly worth following up in that it opens up still another possibility to develop a rational mechanism of scoring that is not inconsistent with the latest knowledge about adhesion and scoring. For the time being, however, as Fein himself notes, it would not appear feasible to derive from his mechanism any sufficiently quantitative predictions, such as about the levels of scoring temperatures for given combinations of steels and nonadditive mineral oils. For this predictive purpose, it is a pity that Tabor and others have devoted much less attention to the junction growth of steels than to that of other metals and alloys.

Any further study of Fein's mechanism should preferably include the possible effect of work-hardening of various kinds of steel on junction growth, and thereby on scoring temperature. There are at least indications that with comparatively soft steels the scoring temperatures tend to be lower than with hard steels that are lubricated with the same non-additive mineral oil. (See the author's reply to Mr. Dyson's written discussion and also fig. 3.7 of the paper of J. P. O'Donoghue and A. Cameron (ref. 64).)

In his first discussion, Dr. Fein plots, on a common basis, the critical temperatures calculated by him on the one hand for the four-ball machine and on the other for gear tests of M. Ibrahim and A. Cameron. He ascribes the quite reasonable strictness of the correlation to the fact that in these gear tests the speeds and blank temperatures were low, like in the four-ball machine.

The author, however, in his discussion on W. C. Pike and D. T. Spillman's recent ASLE Paper 69-LC-4, "Effects of Seizure-Delay on Transition Temperatures in the Four-Ball Machine," has pointed out that the usual method of calculating the scoring temperatures in the four-ball machine needs considerable refinement. Charts for performing such refined calculations had already been worked out by the author prior to World War II but went astray owing to the war conditions in his country. These charts are now being reconstructed, and it is hoped that they can soon be made available to all those interested.

It is thought that only after the data in Dr. Fein's figure have been recalculated by means of the refinements concerned can it be judged to what extent the suggested correlation can be upheld.

Messrs. Beerbower and Tao may well be right in claiming that, in order to satisfactorily explain the constancy of scoring temperature, scoring

mechanisms different from those suggested by the author may be needed. However, notwithstanding their painstaking efforts in predicting quite a variety of basic characteristics of the oils concerned, it is not thought that they have succeeded in discrediting all of the author's mechanisms.

For instance, they admit that the accuracy of their tabulated estimates of the vapor pressures at the scoring temperatures of the various oils leaves something to be desired. Moreover, the author fails to see why the mechanism of collapse of the lubricant film through cavitation by evaporation at or near the outlet of this film should stand or fall with the equality of the various vapor pressures.

Soon after the author set up the correlation of scoring temperature with viscosity grade (see fig. 1), he struck upon the conjecture that cavitation by evaporation plays a physically significant part in scoring. In fact, at that time he found out, like Beerbower and Tao, that either limit to the uncertainty range of figure 1 corresponds with a nearly constant viscosity at atmospheric pressure. Further, at least within any given truly homologous series of liquids, like that of the *n*-paraffins, any given viscosity seems to correspond with a certain vapor pressure, irrespective of the particular member considered within the series, provided that the temperatures of the individual members are so adjusted as to yield the same viscosity.

In 1965 a few exploratory experiments for verifying the reality of the effect of evaporation, or say of the volatility, of nonadditive mineral oils were carried out on the IAE Standard Spur Gear Rig by the author's then chief assistant, F. H. Theyse. In these tests two lubricants were compared, one being a commercially available nonadditive mineral oil and the other being a blend of this oil with a small amount of a much more volatile hydrocarbon fuel. As expected, the scoring performance of these two lubricants proved to be significantly different. An overall study of certain thermal phenomena characteristics of evaporation in the lubricant film could be made by means of highly sensitive and quickly responding temperature probes that were inserted in some of the gear teeth. This study was indeed indicative of the importance of evaporation in the lubricant films concerned. Unfortunately, owing to lack of staff and the higher priority to be given other work, Theyse's highly intricate experiments could not so far be continued.

The author fails to see the adequacy of Beerbower and Tao's line of reasoning for explaining the constancy of scoring viscosity on the basis of hydrodynamic theory. Both in the classical theory of hydrodynamic lubrication, where the rubbing surfaces are assumed to be perfectly rigid, and in the elastohydrodynamic theory, the representative viscosity is that at the inlet temperature of film rather than that at the maximum surface temperature. But constancy of the latter viscosity does not involve constancy of the former. Moreover, in both of the above-



mentioned hydrodynamic theories, the effects of representative viscosity are closely bound up with those of the entraining velocity, which may be put equal to the sum of the tangential velocities of the two rubbing surfaces relative to the film. In fact, in the formulas resulting from these theories, the representative viscosity appears exclusively in the form of a factor in its product with the entraining velocity. It follows that whenever the entraining velocity at incipient scoring varies from one test to another, as was the case in the experiments concerned, constancy of the scoring viscosity does not ensure constancy of this product, or say of hydrodynamic load capacity.

Mr. Dyson's first comment about the inapplicability of Korovchin-ski's theory of the "thermal bulge" in cases where Péclet number,  $VB/a$ , is not unusually small is correct. Additional observations about this theory may be found in the present, revised text of the paper. The author wishes to apologize for not having submitted the revision concerned long enough before the symposium so that it could be incorporated in the preliminary proceedings on which Dyson based the present comment.

As regards the second and third comment of Mr. Dyson's, the following observations are put forward:

Charts similar to figure 10 were plotted by the author soon after the German translation (ref. 82) of Genkin, Kuz'min, and Misharin's series of articles was published. These "elastohydrodynamic" charts initially induced the author to a conjecture similar to that of Dyson; namely, that scoring temperature in Genkin and associates' disc experiments might perhaps have been influenced by elastohydrodynamic effects.

A more thorough perusal of the abovementioned series of articles, however, led the author to the conclusion that at least in the great majority of the so-called "scoring" tests performed by Genkin and associates, insofar as these related to the two hardest of their three disc steels, the surface damage cannot have been "gross damage" as specified in the aforementioned definition of scoring of the Institution of Mechanical Engineers. In fact, as follows among other things from their description of transient frictional phenomena induced at the onset of the actual surface damage, this damage was not of the self-aggravating type that is meant both in the above definition and in the author's postulate about the constancy of scoring temperature.

Thus the author could not escape the conclusion that Genkin and associates' disc tests on their two hardest steels do not relate to true scoring and, therefore, do not lend themselves to verifying the author's postulate.

Even so, however, the author wishes to emphasize that the present tests yet prove valuable for another important purpose, that is, for establishing correlations between the transition from full to partial elastohydrodynamic lubrication and the various elastohydrodynamically influential variables.



It now remains to consider those of Genkin and associates' disc tests that were performed with their softest steel, Steel 45, of 350 Brinell hardness. They stressed that this steel had a much greater welding tendency than their two harder steels, and that this would have reflected itself in the results obtained with both test oils, Oil MC-20 and Oil MK-8.

Now, for reasons that can be derived from the present paper, the author will, in the following considerations, not include those tests (as he has neither done with the two hardest steels) where load was held constant and peripheral velocities were varied, either up or down, till friction rose noticeably. It is thought that in the remaining tests with the present steel, where peripheral velocities were held constant and load was increased, a similar rise in friction may well have been indicative of true, self-aggravating scoring. In other words, these particular tests, in contrast to all the others, may well lend themselves to verifying the postulate about the constancy of scoring temperature.

Among the 45 tests performed on the present steel with the more viscous of the two test oils, MC-20, there were 40 in each of which the peripheral velocities were held constant, although they differed from one such test to another. In 28 of the latter tests the specific sliding amounted to 1.5, while in the 12 others it amounted to 2.63. In all of these 40 tests the sliding velocity covered the fairly wide range from 2.5 to 13.0 m/sec.

Now, in the first series, comprising the abovementioned 28 tests, the critical maximum surface temperatures estimates by Genkin and associates fell within the range of  $205 \pm 15^\circ \text{C}$ . The width of this range does not exceed the scatter in scoring performance in even the most strictly controlled scoring tests. Moreover, one should account also for uncertainties in the assessment of the scoring temperatures from the scoring performance. Therefore, the narrowness of this width might, if taken all by itself, be considered to support the postulate about the constancy of scoring temperature rather than to be significantly indicative of elastohydrodynamic influences upon this temperature.

Further, in the second series, comprising the abovementioned 12 tests, the critical maximum surface temperatures estimated fell within the range of  $175 \pm 20^\circ \text{C}$ . Notwithstanding this range is somewhat wider than the foregoing one, it too does, again if taken all by itself, support the postulate rather than any significant elastohydrodynamic influences.

But if the two test series are taken together it is seen that, since they do not overlap each other to any worthwhile statistically significant extent, they do not seem to support the postulate, at least not at first sight. It therefore remains to be investigated whether the postulate is really invalidated, if only by the comparatively small average difference of  $205 - 175 = 30^\circ \text{C}$  between the two above levels of scoring temperature, 205 and  $175^\circ \text{C}$  for the sliding ratios of 1.5 and 2.63, respectively.

It is thought, however, that the present difference may well fall within certain systematic inaccuracies in Genkin and associates' estimates of the various critical maximum surface temperatures or, say, of the scoring temperatures. In fact, there still is considerable room for various refinements in their estimating procedures, and these might well result in a corrected difference so small that a common, and comparatively narrow range might be assigned to the scoring temperatures in all of the  $28+12=40$  tests concerned.

Perhaps the most important such refinement, which may well result in an appreciable reduction of the present difference in Genkin and associates' two scoring temperature levels, relates to the inequality of the bulk temperatures,  $T_{b,1}$  and  $T_{b,2}$ , of the two mating discs at incipient scoring. In fact, there must have been such a thermal asymmetry, the two discs having unequal diameters and their peripheral velocities being also unequal. But they measured the bulk temperature of only one of the two discs and took this to have been common to both discs, thus assuming,  $T_{b,1}=T_{b,2}=T_b$  (see eq. (1a), in conjunction with eq. (1b)). The present refinement consists in applying equation (4b), thereby well accounting for the aforementioned difference between  $T_{b,1}$  and  $T_{b,2}$ . Unfortunately the data available from Genkin and associates' paper did not enable the author to work out the present, or in fact any other worthwhile, refinement. The same is true of the remaining tests with the soft Steel 45, those with the least viscous oil, MK-8.

A concluding comment on Mr. Dyson's discussion relates to Leach and Kelley's disc experiments on scoring (ref. 22). Dyson states that, whereas Leach and Kelley themselves concluded that film thickness calculated from the elastohydrodynamic theory had no discernible effect on scoring temperature, there might yet have been such an effect. However, since the scatter in Leach and Kelley's scoring temperatures around their average level cannot very well have exceeded the uncertainties in their assessment, it is doubted whether the effect suggested can, from their tests, ever be proved in any statistically significant manner.

Mr. Ku has excellently succeeded, when substituting in the author's unforeseen absence from the present symposium, in bringing out the salient points of the paper and, through posing certain broad questions, in facilitating a better understanding and thereby in stimulating the discussion. The author will now proceed to answer his questions.

First of all, the author agrees that the conception of a critical maximum surface temperature as a significant criterion for the inception of scoring is in essence more important than the validity of the postulated constancy of this temperature. That is, even if the postulate should eventually prove to have too limited an applicability, its basic conception of scoring temperature as a criterion, then to be considered variable in some way or another, would still be valuable. Of course, ever since the author



evolved the postulate, he has constantly endeavored to verify it according as he could lay his hands on relevant evidence from various sources. This attitude may also explain the length and thoroughness of the present paper.

To begin with, an important exception to the postulate was advanced by the author long ago, that is, the inapplicability of constancy of scoring temperature to at least the great majority of EP oils. Further, it is admitted that insight and data now available still leave considerable room for doubts about the applicability of this constancy, even for the rather limited range of oils of the nonadditive mineral type for which it was originally proposed and ever since systematically put to test by the author.

On the other hand, as is evidenced by the author's emphasis on the need of refinements in assessments of scoring temperatures (see the paper, particularly the appendices), the author still raises at least equally severe doubts about the correctness of previous claims by others that even for nonadditive mineral oils the postulate would break down in their ranges of operating conditions. It is hoped that the availability of the present refinements will stimulate others, as it has done the author, to find out more about the real limitations to the applicability of the postulate with nonadditive mineral oils. Accordingly the author will gladly study also the evidence about such limitations that Mr. Ku has announced but not yet been able to publish.

In fact, during the author's prolonged studies on such limitations, he has repeatedly encountered pieces of evidence, mainly from literature, that at first sight looked promising in the way of providing definite information about the real limitations sought. But unfortunately he has not so far succeeded in gathering any sufficiently definite and clearcut information.

The author therefore expresses the hope that where he may have failed, others may succeed in the not too distant future. It is further hoped that the real limitations ultimately to be found can be cast into forms that will prove both workable and meaningful not only to oil technologists and metallurgists, but also to designers of gears and other machine elements having also counterformal rubbing surfaces.

In conclusion, the author expresses his gratitude toward the discussors for the great pains taken in offering their most stimulating comments.

#### NOMENCLATURE

<i>a</i>	thermal diffusivity, defined by $a = k/\rho c$
<i>A</i>	area
<i>b</i>	thermal contact coefficient, defined by $b = (k\rho c)^{1/2}$
<i>B</i>	width of tooth face, or length of conjunction zone
<i>c</i>	specific heat per unit mass



$C$	symbol for functional relationship
$E$	Young's modulus of elasticity, or complete elliptic integral of the second kind
$E_r$	reduced modulus of elasticity, defined by $1/E_r = \frac{1}{2}[(1 - \nu_1^2)/E_1 + (1 - \nu_2^2)/E_2]$
$f$	coefficient of friction
$F$	symbol for functional relationship
$h$	coefficient of heat transfer, also called, film coefficient
$h_{\min}$	minimum film thickness
$k$	heat conductivity or, as an argument of elliptic integrals, $k = (x/w)^{1/2}$
$K$	Korovchinski's thermal-bulging parameter (see eqs. (5a) or (5b))
$n$	ratio, defined by $n = b_1\sqrt{V_1}/b_2\sqrt{V_2}$
$N_M$	frictional heat generated in the meshing zone of a pair of gears
$p$	local film pressure
$Pé$	Péclet number, defined by $Pé = Vw/a$
$Q$	heat generated per unit time
$Q_v$	heat generated per unit time and per unit volume
$q$	heat flux
$q_{\max}$	maximum heat flux
$R$	radius of conformity, defined by $1/R = 1/R_1 + 1/R_2$
$R_c$	thermal constriction resistance
$R_K$	effective radius of conformity, derived from Korovchinski's theory of thermal bulging
$R_{th}$	thermal resistance
$t$	integration variable
$T$	local flash temperature
$T_b$	bulk temperature
$T_{b,s}$	bulk temperature at incipient scoring
$T_c$	conjunction temperature
$T_f$	maximum flash temperature
$T_{f,s}$	maximum flash temperature at incipient scoring
$T_i$	incipient (grazing) contact temperature
$T_{i,\max}$	maximum incipient (grazing) contact temperature
$T_M$	fictitious bulk temperature of the meshing zone of a pair of gears
$u$	velocity of flow of oil in the longitudinal direction of the film
$V$	tangential velocity of rubbing surface, with respect to the conjunction zone
$V_s$	sliding velocity
$V_\Sigma$	sum velocity, defined as the sum of the tangential velocities of the two rubbing surfaces, with respect to their conjunction zone

$w$	width of the conjunction zone
$w_{\min}, w_{\max}$	minimum and maximum width of tapered conjunction zone, respectively
$W$	load per unit length of the conjunction zone
$x$	coordinate in the longitudinal direction of an oil film, or in the direction of the width of the conjunction zone
$y$	coordinate in the direction of the normal to a rubbing surface
$z$	coordinate in the direction of the thickness of an oil film
$\alpha$	pressure coefficient of viscosity, defined by $\alpha = (1/\eta) \cdot d\eta/dp$
$\alpha_0$	representative pressure coefficient of viscosity
$\gamma$	heat-partitioning ratio, defined by $\gamma = Q_1/Q_2$
$\Delta T$	temperature drop across a thermal resistance
$\epsilon$	coefficient of linear thermal expansion
$\eta$	dynamic or absolute viscosity
$\eta_0$	representative viscosity
$\nu$	Poisson's ratio
$\xi$	integration variable
$\rho$	density
$\sigma_H$	maximum contact pressure according to Hertz
$\sigma_K$	maximum contact pressure according to Korovchinski
$\tau$	shear stress

## REFERENCES

1. BLOK, H.: Surface Temperatures under Extreme-Pressure Lubricating Conditions (in French). Congr. mondialópetr.le, 2 me Congr. (Paris), vol. 3, 1937, pp. 471-486.
2. BLOK, H.: Measurement of Temperature Flashes on Gear Teeth under Extreme-Pressure Conditions. Proc. Gen. Disc. Lubrication, Inst. Mech. Engrs., pt. 2, 1937, pp. 14-20.
3. BLOK, H.: Theoretical Study of Temperature Rise at Surfaces of Actual Contact under Oiliness Lubricating Conditions. Proc. Gen. Disc. Lubrication, Inst. Mech. Engrs., pt. 2, 1937, pp. 222-235.
4. HOFER, H.: Die zulässige Zahnradbeanspruchung und ihre Berechnungsweise im Werkzeugmaschinenbau. Werkstattstechnik, vol. 25, no. 5, 1931, pp. 128-131.
5. HOFER, H.: Vorläufige und verbesserte Zahnradberechnungsarten. Automobil-techn. Zeitschr., vol. 48, 1946, pp. 4-6, and 24-25.
6. ALMEN, J. O.: Factors Influencing the Durability of Spiral-Bevel Gears for Automobiles. Autom. Ind., vol. 73, 1935, pp. 662-668 and 696-701.
7. ALMEN, J. O.: Dimensional Value of Lubricants in Gear Design. SAE J., vol. 50, 1942, pp. 373-380.
8. ALMEN, J. O.: Surface Deterioration of Gear Teeth. Mechanical Wear, J. T. Burwell, Jr., ed., Am. Soc. Metals, ch. 12, 1948, pp. 229-288.
9. BLOK, H.: Seizure Delay Method for Determining the Protection Against Seizure Afforded by Extreme Pressure Lubricants." SAE J., vol. 44, 1939, pp. 193-201 and 220.
10. DUDLEY, D. W.: Practical Gear Design. McGraw-Hill Book Co., Inc., 1954. (See in particular, section 54, pp. 141-144.)

11. KUDRYAVTSEV, V. N.: *Zubchatiye Peredachi*. Gosudarstvennoe Nauchno Tekhnicheskoe Izdatelstvo, Mashinostroitelnoi Literaturi, Moscow and Leningrad, 1957. (See in particular, section 24, pp. 182-186.)
12. LEMANSKI, A. J.: Gear Scoring Design Guide for Aerospace Spur and Helical Power Gears. Proposed Tentative AGMA Information Sheet, AGMA 217.01; Am. Gear Mfrs. Assoc., Washington, D.C.
13. DYSON, A.; AND NAYLOR, H.: Application of the Flash Temperature Concept to Cam and Tappet Wear Problems. *Proc. Inst. Mech. Engrs., Autom. Div.*, 1961, p. 255.
14. NAYLOR, H.: Cam and Friction Drives. *Proc. Third Intern. Conf. Lubrication and Wear, Inst. Mech. Engrs.*, 1967, pp. 66-76.
15. BORSOFF, V. N.: On the Mechanism of Gear Lubrication. *Trans. ASME, J. Basic Engrg.*, vol. 80D, 1959, pp. 79-93.
16. NIEMANN, G.; AND LECHNER, G.: The Measurement of Surface Temperatures on Gear Teeth. *Trans. ASME, J. Basic Engrg.*, vol. 87D, 1965, pp. 641-651.
17. MENG, V. V.: Investigation of Steel Seizure on a Roller Testing Machine. (in Russian). *Friction and Wear in Machinery*, vol. 14, 1960. English translation by ASME, pp. 202-217.
18. FLECK, W.: Beitrag zur Klärung der Fresstragfähigkeit bei Wälzgleiten. *Maschinenbautechnik (Getriebetechnik)*, vol. 12, 1963, pp. 209-216 and 438-444.
19. EMMONS, H.: Theory and Application of Extended Surface Thermocouples. *J. Franklin Inst.*, vol. 229, 1940, pp. 29-52.
20. SHU, H. H. H.; GAYLORD, E. W.; AND HUGHES, F. W.: The Relation between the Rubbing Interface Temperature Distribution and Dynamic Thermocouple Temperature. *Trans. ASME, J. Basic Engrg.*, vol. 86D, 1964, pp. 417-422.
21. BLOK, H.: Lubrication as a Gear Design Factor. *Proc. Intern. Conf. Gearing, Inst. Mech. Engrs.*, 1958, pp. 144-158.
22. LEACH, E. F.; AND KELLEY, B. W.: Temperature—The Key to Lubricant Capacity. *ASLE Trans.*, vol. 8, 1965, pp. 271-285.
23. KELLEY, B. W.: A New Look at the Scoring Performance of Gears. *SAE Trans.*, vol. 61, 1953, pp. 175-185.
24. KELLEY, B. W.; AND LEMANSKI, A. J.: Lubrication of Involute Gears. *Proc. Third Intern. Conf. Lubrication and Wear, Inst. Mech. Engrs.*, 1967, pp. 173-184.
25. LECHNER, G.: Fresstragfähigkeit der Zahnräder. *ASUG Mitteilungen (Magdeburg, E. Germany)*, vol. 5, no. 1, 1968, pp. 23-25.
26. LING, F. F.; AND MOW, V. C.: Surface Displacement of a Convective Elastic Half-Space under an Arbitrarily Distributed Fast-Moving Heat Source. *Trans. ASME, J. Basic Engrg.*, vol. 87D, 1965, pp. 729-734.
27. KOROVCHINSKI, V. N.: Plane-Contact Problem of Thermoelasticity during Quasi-Stationary Heat Generation on the Contact Surfaces. *Trans. ASME, J. Basic Engrg.*, vol. 87D, 1965, pp. 811-817.
28. LING, F. F.; AND RICE, J. S.: Surface Temperature with Temperature-Dependent Thermal Properties. *ASLE Trans.*, vol. 9, 1966, pp. 195-201.
29. TERAUCHI, Y.; AND MIAYO, Y.: On the Measurement of Temperature Flashes on Spur Gear Teeth. *Bull. Japan Soc. Mech. Engrs.*, vol. 7, no. 26, 1964, pp. 444-451; vol. 8, no. 29, 1965, pp. 109-119.
30. BLOK, H.: The Flash Temperature Concept. *Wear*, vol. 6, 1963, pp. 483-494.
31. ALLEN, D. N. DE G.: A Suggested Approach to Finite-Difference Representation of Differential Equations, with an Application to Determine Temperature-Distributions near a Sliding Contact. *J. Mech. and Appl. Math.*, vol. 15, pt. 1, 1962, pp. 11-13.
32. CAMERON, A.; GORDON, A. N.; AND SYMM, G. T.: Contact Temperatures in Roll-



- ing/Sliding Surfaces. *Proc. Roy. Soc. (London)*, vol. A286, 1965, pp. 45-61.
33. SYMM, G. T.: Surface Temperatures of Two Rubbing Bodies. *J. Mech. and Appl. Math.*, vol. 20, pt. 3, 1967, pp. 381-391.
34. OSTERLE, F.; CARNES, A.; AND SAIBEL, E.: On the Solution of the Reynolds Equation for Slider-Bearing Lubrication—IV. Effect of Temperature on the Viscosity. *Trans. ASME*, vol. 75, 1953, pp. 1117-1123.
35. LING, F. F.; AND SIMKINS, T. E.: Measurement of Pointwise Junction Condition of Temperature at the Interface of Two Bodies in Sliding Contact. *Trans. ASME, J. Basic Engrg.*, vol. 85D, 1963, pp. 481-487.
36. MOORE, C. J.; ATKINS, H.; AND BLUM, H. A.: Subject Classification Bibliography for Thermal Contact Resistance Studies. ASME Paper 68-WA/HT-18, 1968.
37. KELLEY, B. W.: The Importance of Surface Temperature to Surface Damage. *Handbook of Mechanical Wear*, C. Lipson and L. V. Calwell, eds., Univ. Michigan Press, Ann Arbor, Mich., 1961, chap. 8, pp. 151-172.
38. GRUBIN, A. N.: Investigation of the Contact of Machine Components, (in Russian). *Minist. Instit. Tekh., Mashin. Tssnit. Nauchno-Issledovetel. Inst. Teknol. i Maskinostroeniya, Mashin Lit. (Moscow)*, vol. 30, 1949, K. F. Ketova, ed. English translation issued in 1956 by Dept. Sci. Ind. Res. (London), Translation 337.
39. CROOK, A. W.: The Lubrication of Rollers. III. *Phil. Trans. Roy. Soc. (London)*, vol. A254, 1961, p. 237-258.
40. DOWSON, D.; AND WHITTAKER, A. V.: A Numerical Procedure for the Solution of the Elastohydrodynamic Problem of Rolling and Sliding Contacts Lubricated by a Newtonian Fluid. *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1965-66, pp. 27-41.
41. CHENG, H. S.; AND STERNLICHT, B.: A Numerical Solution for the Pressure, Temperature, and Film Thickness between Two Infinitely Long, Lubricated Rolling and Sliding Surfaces, under Heavy Loads. *Trans. ASME, J. Basic Engrg.*, vol. 87D, 1965, pp. 695-704.
42. MANTON, S. M.; O'DONOGHUE, J. P.; AND CAMERON, A.: Temperatures at Lubricated Rolling/Sliding Contacts. *Proc. Inst. Mech. Engrs.*, vol. 182, pt. 1, 1968.
43. DOWSON, D.; AND HIGGINSON, G. R.: *Elastohydrodynamic Lubrication. The Fundamentals of Roller and Gear Lubrication.* Pergamon Press, Oxford, England, 1966.
44. VAN DER HELD, E. F. M.: Solving Contact Problems through Series Expansions with Terms consisting of Gaussian Error Integrals, (in Dutch). *Warmte-Techniek (Holland)*, no. 12, 1940, pp. 3-16.
45. BLOK, H.: Gear Lubricant—A Constructional Gear Material. *De Ingenieur (Holland)*, vol. 63, no. 39, 1951, pp. 0.53-0.64.
46. *Members of MIT Staff: Electric Circuits.* John Wiley & Sons, New York, N.Y., 1948.
47. BLACKWELL, W. A.: *Mathematical Modeling of Physical Networks.* MacMillan Co., New York, N.Y., 1968.
48. SLEPIAN, P.: *Mathematical Foundations of Network Analysis.* Springer, Berlin, 1968.
49. WELLAUER, E. J.: The Thermal Problem in Enclosed Gear Drives. *Am. Gear Mfrs. Assoc., Semiannual Meeting*, October 1951.
50. OHLENDORF, H.: Verlustleistung und Erwärmung von Stirnradgetrieben. Doctorate thesis, Technische Hochschule, Munich, W. Germany, 1958. See also extensive abstract in *Stirnradgetriebe, Zahnradreibung, Verlustleistung und Erwärmung*, W. Richter, ed., F. Vieweg & Sohn, Braunschweig, W. Germany, 1964.
51. DUSEV, I. I.: Analytical Investigation of Hypoid Gears with Linear Contact of

- the Touching Surfaces, (in Russian). Trudy Novocherk, Politekh. In-ta, vol. 126, 1961, pp. 73-94.
52. POWELL, D. L.; AND HOYL, I. R.: Automotive Hypoid Gear Loading and Sliding Relationships. Proc. Symp. Gear Lubrication, Inst. Petroleum (London), 1964, pp. II. 1-7.
  53. COLEMAN, W.: Bevel and Hypoid Gear Surface Durability: Pitting and Scuffing. Proc. Conf. Lubrication and Wear, Inst. Mech. Engrs., 1967, pp. 191-204.
  54. FLETCHER, A. G.; AND MCWATERS, J. F.: Calculation of Contact Conditions in Hypoid Gears. Report No. 202, Natl. Engrg. Lab., East Kilbride, Scotland, 1965.
  55. SHURYGIN, YU. I.: Geometrical Relationships of the Initial Surfaces of a Hypoid Transmission, (in Russian). Maschinostroeniya, no. 4, 1965, pp. 22-27.
  56. KECK, K. F.: Einfluss der Zahnform auf die Gleit- und Verschleiss-verhältnisse von Hypoid-Kegelrädern. Konstruktion, vol. 18, no. 4, 1966, pp. 163-166.
  57. DYSON, A.: A General Theory of the Kinematics and Geometry of Gears in Three Dimensions. Oxford Univ. Press, 1968.
  58. Memorandum of Definitions, Symbols, and Units. Proc. Intern. Conf. Lubrication and Wear, Inst. Mech. Engrs., 1957, pp. 1-5.
  59. BEANE, G. A., IV; AND LAWLER, C. W.: Lubricant Evaluation with the WADC High-Temperature Gear Machine. Proc. Symp. Gear Lubrication, Inst. Petroleum (London), 1966, pp. I. 37-53.
  60. VINOGRADOV, G. V.; ARKHARARIVA, A.; AND PETROV, A. A.: Anti-Wear and Anti-Friction Properties of Hydrocarbons under Heavy Loads. Wear, vol. 4, 1961, pp. 274-291.
  61. VINOGRADOV, G. V.; KOREPOVA, I. V.; PODOLSKY, YU. YA.; AND PAVLOVSKAYA, N. T.: Effect of Oxidation on Boundary Friction of Steel in Hydrocarbon Media and Critical Friction Duties with Cold and Hot Seizure. Trans. ASME, J. Basic Engrg., vol. 87D, 1965, pp. 741-746.
  62. ASKWITH, T. C.; CROUCH, R. F.; AND CAMERON, A.: The Relationship of Molecular Chain Length of Lubricants and Additives for Optimum Effect and a Theory of Scuffing. Proc. Symp. Gear Lubrication, Inst. Petroleum (London), 1966, pp. II. 16-34.
  63. ASKWITH, T. C.; CROUCH, R. F.; AND CAMERON, A.: Chain Length of Additives in Relation to Lubricants in Thin Film and Boundary Lubrication. Proc. Roy. Soc. (London), vol. A291, 1966, pp. 500-519.
  64. O'DONOGHUE, J. P.; AND CAMERON, A.: Temperature at Scuffing. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965-6, pp. 85-94.
  65. MATVEEVSKY, R. M.: The Critical Temperature of Oil in Point and Line Contact Machines. Trans. ASME, J. Basic Engrg., vol. 87D, 1965, pp. 754-759.
  66. NIEMANN, G.; AND ASSMANN, H.: Experience with the FZG Apparatus for Testing Gear Oils. Proc. Gear Lubrication Symp., Inst. Petroleum (London), 1964.
  67. EVANS, L. S.: Discussion. Proc. Inst. Petroleum (London), vol. 38, 1952, p. 689.
  68. BORSOFF, V. N.; AND GODET, M.: A Scoring Factor for Gears. ASLE Trans., vol. 6, 1963, pp. 147-153.
  69. PETRUSEVICH, A. I.: The Basic Conclusions from the Elastohydrodynamic Theory of Lubrication, (in Russian). Izv. Akad. Nauk SSSR, Odt. Tekhn. Nauk, no. 2, p. 209. English translation SST-138, BST-480, Science Transl. Service, Univ. Alabama.
  70. BORSOFF, V. N.; ACCINELLI, J. B.; AND CATTANEO, A. G.: The Effect of Oil Viscosity on the Power-Transmitting Capacity of Spur Gears. Trans. ASME, vol. 73, 1951, pp. 687-696.
  71. BORSOFF, V. N.; CROOK, D. L.; AND OTVOS, J. W.: Tracer Technique for Studying Gear Wear. Nucleonics, vol. 10, no. 10, 1952, pp. 67-69.



72. BORSOFF, V. N.: Wear Studies with Radioactive Gears. *Lubrication Engrg.*, vol. 12, 1956, pp. 24-28.
73. IBRAHIM, M.; AND CAMERON, A.: Oil Film Thickness and the Mechanism of Scuffing in Gear Teeth. *Proc. Lubrication and Wear Conv., Inst. Mech. Engrs.*, 1963, pp. 228-238.
74. BOERLAGE, G. D.; AND BLOK, H.: The Four-Ball Top for Testing the Boundary Lubricating Properties of Oils under High Mean Pressures. *Engineering*, vol. 144, no. 3729, 1937, pp. 1-2.
75. O'DONOGHUE, J. P.; MANTON, S. M.; AND ASKWITH, T. C.: An Experimental Determination of the Temperature at Scuffing. *Proc. Tribology Convention, Inst. Mech. Engrs.*, 1967, pp. 20-25.
76. LANCASTER, J. K.: The Formation of Surface Films at the Transitions between Mild and Severe Metallic Wear. *Proc. Roy. Soc. (London)*, vol. A273, pp. 466-483.
77. CHENG, H. S.; AND ORCUTT, F. K.: A Correlation between Theoretical and Experimental Results on the Elasto-Hydrodynamic Lubrication of Rolling and Sliding Contacts. *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1965, pp. 158-168.
78. KELLEY, B. W.; AND LEMANSKI, A. J.: Lubrication of Involute Gearing. *Proc. Inst. Mech. Engrs.*, vol. 182, pt. 3A, 1967-8, pp. 173-184.
79. KELLEY, B. W.: Lubrication of Concentrated Contacts—The Practical Problem. This symposium.
80. RUSSELL, J. A.; CAMPBELL, W. E.; BURTON, R. A.; AND KU, P. M.: Boundary Lubrication Behavior of Organic Films at Low Temperatures. *ASLE Trans.*, vol. 8, 1965, pp. 48-58.
81. KU, P. M.: The Effects of Oxide and Organic Films on Sliding Friction. *Fundamental Phenomena in the Material Sciences*, vol. 2, Surface Phenomena, L. J. Bonis and H. H. Hauser, eds., Plenum Press, 1966, pp. 41-55.
82. GENKIN, M. D.; KUZ'MIN, N. F.; AND MISHARIN, Y. U.: Voprosi zaedeniya Zuchatikh Koles (Investigation into the Wear of Gears). *Izdatelstvo Akademiyi Nauk SSSR*. (Moscow), 1959, chap. 5. German translation in *Maschinenbau und Fertigungstechnik der U.D.S.S.R.*, 1961, vol. 3, pp. 248-267.
83. WATSON, H. J.: Testing of Marine Main-Propulsion-Gear Lubricants in Disc Machines. *Proc. Conf. Lubrication and Wear, Inst. Mech. Engrs.*, 1957, paper 68.
84. DE GRUCHY, V. J.; AND HARRISON, P. W.: Development of an Edge-Type Disc Machine and Preliminary Studies of Various Gear Material Lubricant Combinations. *Proc. Lubrication and Wear Conv., Inst. Mech. Engrs.*, 1963, pp. 160-180.
85. KU, P. M.; AND BABER, B. B.: The Effect of Lubricants on Gear Tooth Scuffing. *ASLE Trans.*, vol. 2, 1959, pp. 184-194.
86. NIEMANN, G.; AND LECHNER, G.: Die Fress-Grenzlast bei Stirnrädern aus Stahl. *Erdöl und Kohle, Erdgas, Petrochemie*, vol. 20, no. 2, 1967, pp. 96-106.
87. PADMORE, E. L.; AND RUSHTON, S. G.: Effects of Loading Procedure on Surface Failure in Spur Gears. *Proc. Inst. Mech. Engrs.*, vol. 179, pt. 3J, 1964-65, p. 278.
88. FEIN, R. S.: Transition Temperatures With Four Ball Machine. *ASLE Trans.*, vol. 3, 1960, pp. 34-39.
89. FEIN, R. S.: Operating Procedure Effects on Critical Temperatures. *ASLE Trans.*, vol. 10, 1967, pp. 373-385.
90. FEIN, R. S.: Effects of Lubricants on Transition Temperatures. *ASLE Trans.*, vol. 8, 1965, pp. 59-68.
91. FEIN, R. S.: Operating Procedure Effects on Critical Temperatures. *ASLE Trans.*, vol. 10, 1967, pp. 373-385.



92. TAO, F. F.: The Role of Diffusion in Corrosive Wear. ASLE Trans., vol. 11, 1968, pp. 121-130.
93. KELLER, D. W.; AND ALDRICH, R. G.: Adhesion of Metallic Bodies Initiated by Physical Contact. J. Adhesion, vol. 1, 1969, pp. 142-156.
94. BUCKLEY, D. H.: Influence of Chemisorbed Films on Adhesion and Friction of Clean Iron. NASA TND-4775, September 1968.
95. TABOR, D.: Junction Growth in Metallic Friction: The Role of Combined Stresses and Surface Contamination. Proc. Roy. Soc. (London), vol. A251, 1959, pp. 378-393.
96. FEIN, R. S.; AND KREUZ, K. L.: Discussion on Boundary Lubrication. Interdisciplinary Approach to Friction and Wear, NASA SP-181, 1968, pp. 538-376.
97. ZARETSKY, E. V.; AND ANDERSON, W. J.: Effect of Materials—General Background. This symposium.

# The Structure of Liquids

**H. EYRING**

**University of Utah  
Salt Lake City, Utah**

**M. S. JHON**

**University of Virginia  
Charlottesville, Virginia**

In this paper, early models of the liquid state are briefly reviewed, and the present status of the significant structure theory is presented. For simple substances such as argon, this theory provides a model without adjustable parameters.

For more complicated molecules such as lubricants, a knowledge of certain properties leads to a prediction of the remaining properties. The calculation of such properties is discussed in conjunction with the research needed to understand lubricants. The properties discussed are (1) thermodynamic properties, (2) ordinary and bulk viscosity, (3) thermal conductivity, (4) sound velocity, and (5) surface tension.

A FLUID WHICH, if placed in a closed vessel, conforms at once to the shape of the vessel without necessarily filling the whole of its volume is called a liquid. The first property distinguishes it from a solid and the second from a gas. However, because of the characteristics of the liquid state, i.e., the strong interaction of many particles and their state of disorder, theoretical analysis has lagged behind the theories of the gaseous and crystalline states.

The problem involved is either to calculate the thermodynamic, transport, and other properties of the liquids by constructing an acceptable method (refs. 1 to 5) which leads to a mathematical description of the liquid state, or, alternatively, to start from a radial distribution function, or cluster integrals (refs. 1 to 5), and derive a usable distribution function (refs. 6 to 9). Although the formal approach, omitting a model, starts from an exact mathematical formulation, it soon requires questionable approximations if transport or thermodynamic properties are to be calculated.

In this paper, we review various liquid theories, and then we discuss the present status of the generally useful significant liquid structure model as we have developed it. Calculations of some of the thermodynamic and dynamic liquid properties, most of which are related to the subject of this symposium, are presented.

#### SOME EARLY LIQUID MODELS

The model approach is frequently called the "partition function approach," since the model should lead to a mathematical expression (the partition function) to calculate the properties of liquid.

The partition function for a classical liquid is given as follows:

$$A = -kT \ln f_N \quad (1)$$

where

$$f_N = \frac{1}{h^{3N} N!} \int \exp \frac{-H}{kT} dq_N dp_N$$

Here  $A$  is the Helmholtz free energy and  $H$  is the sum of the kinetic and potential energies of the  $N$  molecules expressed in terms of their coordinates  $q_N$  and their momenta  $p_N$ . If the evaluation of the partition function is available, we can calculate all thermodynamic properties with the aid of equation (1) using thermodynamics. Unfortunately the direct evaluation of  $f_N$  is impossibly long. Accordingly we need a more perceptive strategy.

#### Cell Model

Among the models, the simplest is the cell model (ref. 10). Each molecule is assumed to be confined to a singly occupied cell formed by its nearest neighbors; and within its cell, each molecule is assumed to move in a field  $\psi(r)$ .

The neighboring molecules, or the wall molecules, are pictured as forming a "cell" or "wall" in which the central molecule moves. This model was made in an attempt to break up the configurational partition function into a product of identical similar integrals, one for each molecule.

Then the partition function becomes

$$f_N = \lambda^{-3N} \exp (-N\psi(o)/2kT) \cdot V_f^N \quad (2)$$

where

$$\lambda = \frac{h}{(2\pi mkT)^{1/2}}$$

and  $V_f$  is called the free volume and is given by

$$V_f = \int_{\text{cell}} \exp \{ -[\psi(r) - \psi(o)]/kT \} dr \quad (3)$$



Eyring et al. (ref. 11) first used the cell model to correlate the properties of fluids. A further quantitative improvement on the cell theory was based on a realistic potential function developed by Lennard-Jones and Devonshire (refs. 12 and 13). They evaluated  $\psi(r)$  and expressed  $V_f$  in terms of  $\psi(r)$  and  $\psi(o)$ . According to them, the interaction energy between identical simple spherical molecules is given as

$$\psi(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (4)$$

where  $\epsilon$  and  $\sigma$  are energy and distance parameters. This 6-12 L-J potential was used in their calculations.

#### Hole or Free Volume Theory

The density and the apparent number of nearest neighbors ordinarily decrease with melting as well as with increasing temperature. Also, the liquid has an entropy increase upon melting (i.e., liquid argon  $\Delta S_f = 3$  eu).

We can explain these facts by postulating vacant lattice sites or holes if we use a quasi lattice picture of the structure of the liquid. An alternative explanation invokes the presence of a "communal entropy" (ref. 14) in the liquid that is not present in the solid but is present in the gas.

The presence of holes explains part of the entropy increase on melting and provides a basis for understanding the ease of transport mechanism of the liquid. The disorder and entropy, however, are further explained by postulating some communal entropy in the cell model. The concept of Cernuschi and Eyring (ref. 15) has been reviewed by Rowlinson and Curtiss (ref. 16).

Next we consider briefly the free volume model partition function. The volume  $V$  of the liquid is divided into  $L > N$  cells so that  $v = V/N$  is the volume per molecule and  $w = V/L$  is the volume per cell. The coordination number of a particle  $i$  will be decreased because of the holes in the liquid, i.e.,  $Z_i = y_i Z$  where  $y_i$  is some number between zero and unity. Then the partition function for the hole model becomes

$$f_N = \left( \frac{L!}{N!(L-N)!} \right) \frac{(2\pi mkT)^{3N/2}}{h^{3N}} \cdot \exp \left( \frac{-yNZ\phi(o)}{2kT} \right) (V_f)^N \quad (5)$$

where

$$\psi(o) = Z\phi(o)$$

For the evaluation of the partition function, we have to know the dependence of  $v_f$  on  $y$ , and the Bragg-William or quasi-chemical approximations are used. Several authors including Cernuschi and Eyring (ref. 15), Rowlinson and Curtiss (ref. 16), Ono (ref. 17), Peek and Hill (ref. 18), and Henderson (ref. 19) assumed a linear dependency of free volume  $v_f$  on  $y$ .

Henderson's results (ref. 19) are very good and show a significant improvement over the earlier attempts. He used the following linear free volume expression:

$$v_f = yv_f^0 + (1-y)q \quad (6)$$

where  $y$  is chosen to minimize the free energy.

Besides these examples, many other modifications of the free volume theory which considers the nonsphericalization of the cell, more careful calculation of the communal entropy, and consideration of the correlation with neighboring molecules have been developed.

So far, no one has significantly improved upon the previous free volume theories. The cell theory or free volume theory overestimates the solid-like properties. One may say that the cell theory is the theory of a superheated solid, rather than the theory of a liquid.

#### Tunnel Theory

To overcome this defect, Barker (refs. 2 and 20) formulated the tunnel model for liquids which is essentially a one-dimensionally disordered cell theory. The basic idea of the tunnel model is that the molecules of the liquid are imagined to be divided into subsystems consisting of lines of molecules moving one-dimensionally in tunnels whose walls are formed by neighboring lines of molecules.

Assuming that the volume,  $V$ , is divided into  $K$  tunnels, each containing  $M$  molecules, and assuming that the longitudinal and transverse motions in a given tunnel are separable, the partition function of the liquid becomes

$$f_N = \lambda^{-3N} (S(1))^K A_f^N \quad (7)$$

where

$$\lambda = \frac{h}{(2\pi mkT)^{1/2}}$$

$S(1)$  is the one-dimensional configurational integral, and  $A_f$  is the free area for motion across a tunnel.

The calculated results for volume and energy are close to the observed values for liquid argon. The calculated entropy, however, is in poor agreement with experiment. This discrepancy apparently arises because the tunnel theory underestimates the volume of configuration space associated with non-solidlike structures; that is, the assumption that the molecules in different tunnels move independently is questionable. The treatment of correlation effects analogous to that of the cell theory will require further development.

#### SIGNIFICANT LIQUID STRUCTURE THEORY

An extension of the early hole theory by Eyring and his associates (refs. 4 and 5) interprets the space in liquids as mobile fluidized vacancies



which are quite different from the locked in static vacancies in the solid, and which are moved by molecules jumping into a vacancy. Vacancies completely surrounded by molecules confer gas-like properties on some one of the neighbors, while a vacancy completely surrounded by other vacancies is without dynamic effect. The average fraction of filled positions next to a vacancy is  $V_s/V$ , which is also the probability that a vacancy will confer gas-like properties on a molecule. This idea leads us to the modern concept of the "significant liquid structure model," the present status of which will now be discussed briefly.

Using argon as an example of the typical behavior of a normal liquid, we now consider the justification and consequences of our model. In spite of the normal expansion (12 percent) upon melting, the decrease of the coordination number, and a marked increase in fluidity, X-ray diffraction (ref. 21) indicates that nearest neighbor distances are not appreciably changed from that in the solid state.

This can be understood if some of the molecules are replaced by fluidized vacancies that simply increase liquid volume by occupying an average volume equal to an empty lattice site and by destroying the long-range order in the liquid by their variation in average size. The following explanation of the law of rectilinear diameters, which states that the mean density of the liquid and vapor is a slightly sloping straight line over the entire liquid range, leads to a clearer understanding of this model of fluidized vacancies. When a molecule is vaporized without leaving a vacancy, the energy spent is just half of all bonds joining the molecule to its neighbors. However, vaporizing a molecule leaving a vacancy involves the breaking of all bonds to its neighbors. The energy needed to make a vacancy is therefore equal to the energy of vaporization.

Because these fluidized vacancies are moved by neighboring molecules jumping into them, a vacancy changes into translations three degrees of freedom that would have been vibrations. In effect, a vacancy acts like a vapor molecule with not only the effective energy of a vapor molecule but with the entropy as well.

Accordingly the behavior of the molecules in the vapor is mirrored in the liquid as vacancies. We therefore expect to find the same number of dynamic vacancies in 1 cc of liquid as there are molecules in 1 cc of vapor (fig. 1). On this basis, the law of rectilinear diameter can be explained. We assume that the most favorable size of the vacancies is such as to just accommodate a molecule, but that there must be a distribution about this most probable size. An instantaneous photograph would reveal an intimate mixture of solid-like and gas-like degrees of freedom in the liquid. This heterogeneity is fine-grained, since there is no solid-like or gas-like nuclei large enough to form focii that would prevent supercooling or bumping of liquids.

In order that loose gas-like regions can exist in a fluid state without



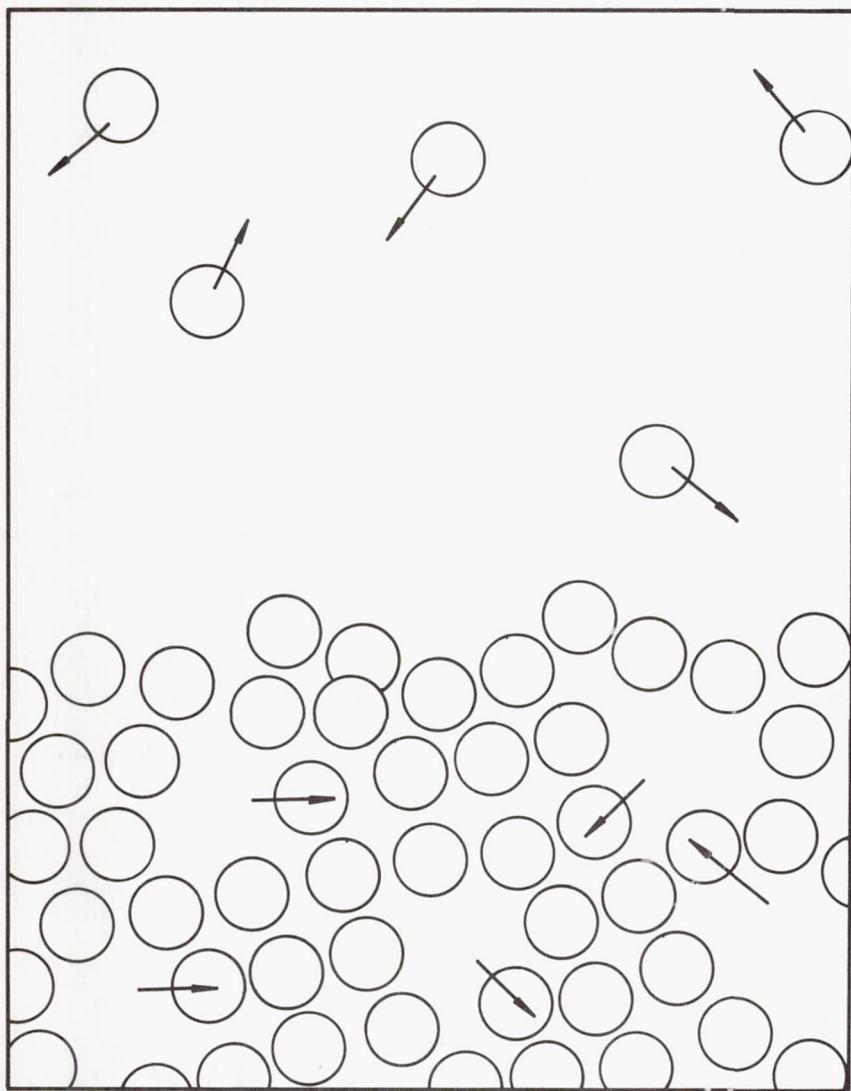


FIGURE 1.—Vacancies in a liquid are the inverse of molecules in a gas. In a liquid, vacancies move among molecules. In a gas, molecules move among vacancies (refs. 5 and 34).

collapse, there must be a local balance between the kinetic energy density tending to make a region expand and the potential energy density tending to make this region collapse. This balance is necessary for a phase to be dynamically stable, and this dynamical stability persists into the supercooled liquid region, where the liquid is no longer thermo-

dynamically stable. The considerations outlined earlier lead to the values

of  $\frac{V_s}{V} \frac{V - V_s}{V_s} = \frac{V - V_s}{V}$  for the mole number of degree of gas-like molecules

and therefore to  $\frac{V_s}{V}$  for the effectively solid-like molecules. Here  $V$  and

$V_s$  are the molar liquid volume and the molar solid-like volume.

According to Walter and Eyring (ref. 22), the specific heat of argon,  $C_v$ , is well represented over the entire liquid range by the equation

$$C_v = 6 \frac{V_s}{V} + 3 \frac{V - V_s}{V} \quad (8)$$

The striking validity of this expression is seen in figure 2.

Other properties such as viscosity, thermal conductivity, and dielectric constants are successfully calculated in analogous fashion, and some of them will be discussed later. Vacancies confer positional degeneracy on the solid-like molecules. The number of additional sites will be the number of vacancies around solid-like molecules ( $n_h = n(V - V_s)/V_s$ ) multiplied by the probability that the molecule has the required energy  $\epsilon$ , to preempt the site from competing neighbors. Here  $\epsilon$  is inversely proportional to  $n_h$ , and directly proportional to the energy of sublimation,  $E_s$ .

In view of the above, the partition function  $f_N$  for a mole of liquid can be expressed as

$$f_N = (f_s)^{N(V_s/V)} (f_g)^{N[(V - V_s)/V]} \quad (9)$$

Here  $N$  is Avogadro's number, and  $f_s$  and  $f_g$  are the partition functions for solid-like molecules, including the degeneracy factors, and for gas-like molecules.

For a simple liquid such as argon, we can write  $f_N$  as follows:

$$f_N = \left\{ \frac{e^{E_s/RT}}{(1 - e^{-\theta/T})^3} \left[ 1 + n \left( \frac{V - V_s}{V_s} \right) e^{-[aE_s V_s / (V - V_s) RT]} \right] \right\}^{N(V_s/V)} \\ \times \left\{ \frac{(2\pi mkT)^{3/2}}{h^3} (V - V_s) \right\}^{N[(V - V_s)/V]} \bigg/ \left( N \frac{V - V_s}{V} \right)! \quad (10)$$

In this derivation, we used the Einstein oscillator as an adequate representation of the solid-like molecule, while the ideal gas partition function is

used for the gas-like molecules. The factor  $\left( N \frac{V - V_s}{V} \right)!$  is used because

the gas-like molecules communally share the extra volume  $(V - V_s)$ .

One of the criticisms raised against this theory is that the partition

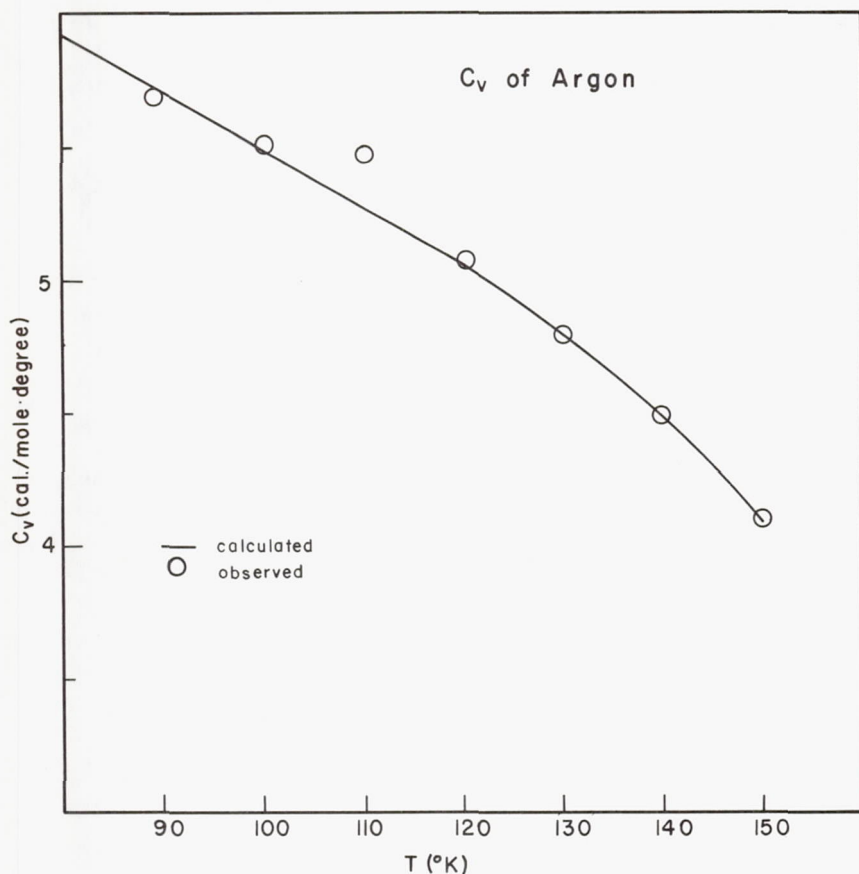


FIGURE 2.—Heat capacity at constant volume for liquid argon. The solid curve represents equation (8), and the circles represent the experimental data (refs. 5 and 22).

function has a number of adjustable parameters: the heat of sublimation,  $E_s$ , the Einstein characteristic temperature,  $\theta$ , the solid-like molar volume,  $V_s$ , and the proportionality constants  $a$  and  $n$ . On the contrary, for simple liquid, the values of  $E_s$ ,  $\theta$ , and  $V_s$  are obtained from the solid state, and so could be calculated theoretically; and the values of  $a$  and  $n$  are now evaluated theoretically.

Near the melting point, the fraction of the neighboring positions to a molecule,  $Z$ , which are empty and therefore available for occupancy, is

$$Z \left( \frac{V_m - V_s}{V_m} \right) = \left( Z \frac{V_s}{V_m} \right) \frac{V_m - V_s}{V_s} = n \frac{V_m - V_s}{V_s} \quad (11)$$

hence



$$n = Z \frac{V_s}{V_m}$$

Here  $V_m$  is the molar liquid volume at the melting point. We next calculate  $a$  (ref. 5).

The average solid-like molecule has kinetic energy equal to  $\frac{3}{2}kT$ . If a molecule is to preempt a neighboring vacancy in addition to its original position, it must introduce more kinetic energy into the vacancy than the remaining neighbors can introduce collectively, since the kinetic energy density is proportional to pressure. The required energy is

$$\epsilon_n = \frac{(n-1)}{Z} \frac{1}{2} \left( \frac{3}{2} kT \right) \quad (12a)$$

Equation (12a) follows because there are  $(n-1)$  competing molecules, each headed toward the vacancy a fraction  $1/Z$  of the time and having a kinetic energy  $T\frac{3}{2}k$  dividing itself equally between the vacancy and the original position, accounting for the factor  $\frac{1}{2}$ . Thus

$$\epsilon_n = \frac{2E_s V_s}{V_m - V_s} = \frac{1}{2} \left( \frac{3}{2} kT \right) \frac{n-1}{Z}$$

here

$$E_m = \frac{3}{2} kT = \frac{(V_m - V_s)}{VN} E_s \quad (12b)$$

Because the temperature does not change in melting, the energy of melting is all potential and has the value  $E_s(V - V_s)/VN$ , but for dynamic stability this must equal the kinetic energy  $\frac{3}{2}kT$ . The theory works without any adjustable parameter for a simple liquid as shown in table 1.

The only notable deviation of a calculated property is the critical pressure. The calculated pressure is improved by the replacement of the ideal gas partition function with one for an imperfect gas (ref. 23) as predicted by the theory. For polyatomic liquids involving restricted rotation and a change in the solid-like structure upon melting, more sophisticated calculations of the parameters are required, or they can be chosen to fit the experiment.

Several efforts (refs. 5 and 24) have been made to relate this model to the free volume model which has been given a statistical mechanical basis by Kirkwood. However, the difficulty of a detailed justification from first principle of our model is related to the difficulty of solving the many-body problems since the fluidized vacancies are moved about by the cooperative action of all its neighbors. Our model has been the most widely applied of the various theories of liquids and should be applied to

TABLE 1.—*Properties of Liquid Argon*

Temp, °K	$P_{\text{calc}}$ , atm	$P_{\text{obs}}$ , atm	$V_{\text{calc}}$ , cc	$V_{\text{obs}}$ , cc	$\Delta S_{\text{calc}}$ , eu	$\Delta S_{\text{obs}}$ , eu
83.96 (mp)	0.6874	0.6739	28.84	28.03	20.07	19.43
87.49 (bp)	1.040	1.000	29.36	28.69	18.90	18.65
97.76	2.883	2.682	31.04	30.15	15.92	.....

	$T_c$ , °K	$P_c$ , atm	$V_c$ , cc
Calc	149.5	52.9	83.5
Obs	150.7	48.0	75.3

lubricants. There seems to be no liquid which cannot be usefully examined using this model.

#### SOME APPLICATIONS

This significant structure model has been applied with good success (ref. 5) to the calculation of thermodynamic properties of the noble gases: the hydrocarbons, the halogens, fused salt and metals, a quantum liquid, the dense gas, hydride compounds, water and electrolyte solutions, and of liquid mixtures, plastic crystals, high polymers (refs. 25 and 26) and so on. Further, the model successfully explains the dynamic liquid properties such as viscosity, thermal conductivity, and dielectric relaxation, as well as interfacial phenomena such as surface tension and physical absorption.

Here we discuss thermodynamic properties, ordinary and bulk viscosity, sound velocity, thermal conductivity and surface tension.

#### Thermodynamic Properties

If the partition function is known, we are able to evaluate thermodynamic properties from the melting point through the critical point and the gas phase for a wide variety of substances (see ch. 4 in ref. 5). The relation between the Helmholtz free energy,  $A$ , and the partition function,  $f$ , is

$$A = -kT \ln f \quad (13)$$

and since  $A$  is thus known as a function of  $V$  and  $T$ , we can calculate all the other thermodynamic properties. Thus

$$S = -\left(\frac{\partial A}{\partial T}\right)_V, \quad E = -T^2 \left(\frac{\partial A/T}{\partial T}\right)_V, \quad p = -\left(\frac{\partial A}{\partial V}\right)_T,$$

$$C_v = T \left(\frac{\partial S}{\partial T}\right)_V = T \left(\frac{\partial^2 A}{\partial V^2}\right)_T$$

with similar expressions for the remaining properties.

## Ordinary Viscosity and the Diffusion Coefficient

According to early derivations (ref. 1) from the rate theory, the viscosity is expressed as follows:

$$\eta = \frac{hN}{V} e^{\Delta G^\ddagger/RT} = \left( \frac{hN}{V} e^{-\Delta S^\ddagger/R} \right) e^{\Delta H^\ddagger/RT} \quad (14)$$

Here  $V$  is the molar volume,  $N$  is Avogadro's number, and  $\Delta G^\ddagger$ ,  $\Delta H^\ddagger$ , and  $\Delta S^\ddagger$  are the free energy of activation, enthalpy of activation, and entropy of activation, respectively.

A major success of this model is that it leads to an expression for viscosity proportional to  $\exp(\Delta H^\ddagger/RT)$ . This Arrhenius form has long been observed to hold approximately for many liquids. One of the defects of this old model, however, is that the viscosity of many liquids changes exponentially with temperature only at low temperatures. Further, the old model is unable to explain properties such as supercooling or the glass transition region. Also, since the fluidity of a liquid (the reciprocal of the viscosity) is proportional to the excess volume ( $V - V_s$ ), i.e., to the number of vacancies introduced, as Batschinski (ref. 27) pointed out, this result should be part of an acceptable theory of the liquid state. The significant structure theory satisfies these requirements.

Efforts have been made by Doolittle (refs. 28 and 29), Cohen and Turnbull (ref. 30), Beuche (ref. 31), McCedo and Litovitz (ref. 32), and Williams, Landel, and Ferry (ref. 33) to develop realistic theories of viscosity.

Let us consider that a fraction,  $X_s$ , of a shear surface is covered by solid-like molecules and the remaining fraction,  $X_g$ , is covered by gas-like molecules.

Then the viscosity,  $\eta$ , which is the ratio of shear stress,  $f$ , to the rate of strain,  $\dot{s}$ , is

$$\eta = \frac{f'}{\dot{s}} = \frac{x_s f'_s + x_g f'_g}{\dot{s}} = x_s \eta_s + x_g \eta_g = \frac{V_s}{V} \eta_s + \frac{V - V_s}{V} \eta_g \quad (15)$$

The term  $\eta_g$  is taken as equal to the gas viscosity obtained from the kinetic theory of gases (ref. 34); i.e.,

$$\eta_g = \frac{2}{3d^2} \left( \frac{mkT}{\pi^3} \right)^{1/2} \quad (16)$$

where  $m$  is the molecular weight, and  $d$  is the diameter of the molecules.

In accord with a generalization (ref. 35) of Eyring's early procedure,  $\eta_s$  is found to have the values:

$$\eta_s = \frac{Nh}{ZK'} \frac{b}{\sqrt{2}} \frac{V}{V_s} \frac{1}{(V - V_s)} \frac{1}{(1 - e^{-\theta/T})} \exp \left( \frac{a' E_s V_s}{RT(V - V_s)} \right) \quad (17)$$



Here  $K'$  and  $a'$  are the transmission coefficient and proportionality constant, respectively, and the other symbols have been defined.

Substituting equations (17) and (16) in (15) gives

$$\eta = \frac{Nh}{ZK'} \frac{b}{\sqrt{2}} \frac{1}{(V - V_s)} \frac{1}{(1 - e^{-\theta/T})} \exp \left( \frac{a'E_s V_s}{(V - V_s)RT} \right) + \frac{V - V_s}{V} \frac{2}{3d^2} \left( \frac{mkT}{\pi^3} \right)^{1/2}$$

This equation is self-consistent since if  $V$  approaches  $V_s$ , then  $\eta$  becomes infinite, and as  $V$  grows compared to  $V_s$ ,  $\eta$  is reduced to  $\eta_0$ . Equation (18) also satisfies Batschinski's relation (ref. 27). In general, this equation is adequate to predict the viscosity. The  $V_s$  to be used is the molar volume of the solid-like structure as it actually exists in the liquid due to the structure change and to thermal expansion or contraction.

The effect of pressure was also taken into account in the successful calculation by Ree et al. (ref. 35). The viscosity equation applies in theory at all pressures and temperature; but  $V$  and  $V_s$  are the values for that pressure and temperature and can be calculated from  $A$  or by correcting known values of  $V$  and  $V_s$  to the new values by using appropriate coefficients of expansion and of compressibility. By using the theory of Eyring, Ree, and Hirai (ref. 36) relating high polymer viscosity to that of small molecules, we can extend equation (18) into the high polymer range. The model for the viscosity of pure liquids is readily adapted to mixtures; the binary liquid partition function (ref. 5) can be used to develop the viscosity equations for mixtures. The viscosity formula being considered is

$$\eta = \frac{hN}{r} \frac{b}{\sqrt{2}} \frac{1}{(1 - e^{-\theta_1/T})^{X_1}} \frac{1}{(1 - e^{-\theta_2/T})^{X_2}} \frac{1}{(V - V_s)} \exp \frac{bE_s V_s}{RT(V - V_s)} + \frac{V - V_s}{V} \left[ \frac{2}{3d_1^2} \left( \frac{m_1 kT}{\pi^3} \right)^{1/2} X_1 + \frac{2}{3d_2^2} \left( \frac{m_2 kT}{\pi^3} \right)^{1/2} X_2 \right] \quad (19)$$

$$r = ZK', \quad E_s = E_{s1}X_1^2 + E_{s2}X_2^2 + 2X_1X_2(E_{s1}E_{s2})^{1/2}(1 + \delta E_s),$$

$$V_s = V_{s1}X_1 + V_{s2}X_2, \quad b = b_1X_1 + b_2X_2, \quad r = r_1X_1 + r_2X_2$$

Suffixes 1 and 2 indicate the quantities of the species 1 and 2 in the mixture, and  $X$  denotes the mole fraction of the  $i$ th species. For the derivation of equation (19), the interaction between the viscosities of the gas-like degree of freedom of species 1 and 2 is neglected. It is further assumed that the parametric values  $r$ ,  $V_s$ ,  $E_s$ , and  $b$  are only concentration dependent, and no extra adjustable parameters are used. Some of the recent unpublished results are shown in tables 2 and 3. The results are satisfactory.

The diffusion coefficient  $D$  also may be calculated by means of the following equation of Eyring and Jhon (ref. 5):

$$D = \frac{kT}{\xi \left( \frac{N}{V} \right)^{1/3} \eta} \quad (20)$$

where  $\xi$  is the effective number of nearest neighbors of a molecule lying in the same plane normal to the direction of shear. For a close-packed structure, this is six.

TABLE 2.—Viscosities of Liquid Carbon Tetrachloride and Benzene

$C_6H_6^a$				$CCl_4^b$			
$\theta_1 = 44.36^\circ \text{ K}$				$\theta_2 = 53.53^\circ \text{ K}$			
$E_{s1} = 9636 \text{ cal/mole}$				$E_{s2} = 8613 \text{ cal/mole}$			
$V_{s1} = 82.83 \text{ cc/mole}$				$V_{s2} = 89.39 \text{ cc/mole}$			
$d_1 = 4.5 \text{ \AA}$				$d_2 = 3.8 \text{ \AA}$			
$b_1 = 9.13 \times 10^{-4}$				$b_2 = 12.17 \times 10^{-4}$			
$\gamma_1 = 4.091$				$\gamma_2 = 2.185$			

T, °K	$\eta_{\text{obs}}, ^\circ$ poise	$\eta_{\text{calc}}$	$\Delta\%$	T, °K	$\eta_{\text{obs}} ^\circ$	$\eta_{\text{calc}}$	$\Delta\%$
285.35	0.00737	0.00754	2.32	293.15	0.01329	0.01375	-3.46
293.15	0.00649	0.00649	0.000	293.15	0.00965	0.00964	0.10
303.15	0.00566	0.00555	1.94	303.15	0.00850	0.00840	1.54
313.15	0.00482	0.00489	-0.30	323.15	0.00651	0.00652	-0.15
333.15	0.00395	0.00400	-1.25	343.15	0.00527	0.00533	-1.14
345.85	0.00350	0.00358	-2.29	353.15	0.00466	0.00484	4.29

<sup>a</sup> The values of  $\theta_1$ ,  $E_{s1}$ ,  $V_{s1}$ , and  $V$  are obtained from  $S$  (ref. 38).

<sup>b</sup> The values of  $\theta_2$ ,  $E_{s2}$ ,  $V_{s2}$ , and  $V$  are obtained from (ref. 39).

<sup>c</sup> Reference 40.

TABLE 3.—Viscosities of Carbon Tetrachloride/Benzene Mixtures

$X_1$	T = 298.15° K	$\delta E_s = 0$	
	$\eta_{\text{obs}}, ^a$ poise	$\eta_{\text{calc}}$	$\Delta\%$
0.0000	0.009	0.009	0.00
0.2266	0.008636	0.008397	2.77
0.4053	0.007731	0.007245	6.29
0.6506	0.007017	0.006587	6.12
0.8326	0.006480	0.006212	4.13
1.0000	0.006042	0.005930	1.85

<sup>a</sup> Reference 41.

### Bulk Viscosity

According to the theory of shear viscosity, the application of the shear stress causes no change in the number of holes present in the liquid state. Hirai and Eyring (ref. 37) assumed that when an external pressure is applied, the number of holes decreases to an equilibrium number for the new pressure in their formulation of the bulk viscosity equation. When pressure is applied to a liquid, there is a readjustment through the loss of fluidized vacancies, and there may be a change in the solid-like structure.

The earlier treatment by Hirai and Eyring can be interpreted in terms of a model of fluidized vacancies, although this has not been carried out as yet. We assume that the fluidized vacancies at any temperature are in equilibrium with the phonons in the liquid quasi lattice. When the pressure is applied, the fluidized vacancies are transformed into phonons which then travel to the outside with the velocity of sound.

Since this velocity is very large, the rate of the liquid's compression is determined only by the shear rate of rearrangement of the liquid molecules, transforming a fluidized vacancy into a phonon. During a decompression, phonons start at the surface and are reabsorbed at some point as fluidized vacancies. This behavior is analogous to the absorption and emission of photons by molecules. In the latter case, the photons travel with the velocity of light until they become absorbed, making an excited molecule.

### Sound Velocity

According to well-known relations, the sound velocity  $C$  obeys the following equation:

$$C = \left( \frac{\gamma'}{\beta_T \rho} \right)^{1/2} = \left( \frac{C_p V}{C_v \beta_T M} \right)^{1/2} \quad (21)$$

Here  $\gamma'$  is the ratio of  $C_p$  to  $C_v$ ,  $\beta_T$  is the isothermal compressibility,  $\rho$  is the density,  $V$  is the molar volume, and  $M$  is the molecular weight. These quantities may be evaluated in terms of the partition function as follows:

$$\begin{aligned} \beta_T^{-1} &= -VkT \left( \frac{\partial^2 \ln f}{\partial V^2} \right)_T \\ C_v &= \left( \frac{\partial E}{\partial T} \right)_V = \left( \frac{\partial kT^2 \frac{\partial \ln f}{\partial T}}{\partial T} \right)_V \\ C_p &= C_v + TV\alpha^2/\beta_T, \end{aligned} \quad (22)$$



here

$$\alpha = -\frac{1}{V} \left[ \frac{\partial^2(-kT \ln f)}{\partial V \partial T} \right]_{T,V} \bigg/ \left[ \frac{\partial^2(-kT \ln f)}{\partial V^2} \right]_T$$

Since the expression for sound velocity  $C$  involves properties which depend upon the second derivatives of the partition function, a theoretical test of this property constitutes a severe test of the model. Some poor models fail to predict these properties. As exemplifying the test, the sound velocity of liquid argon is calculated and is listed in table 4. There is good agreement between theory and experiment.

#### Thermal Conductivity

Significant structure theory also forms the basis for an interpretation of thermal conductivity. Thus we can write

$$K = \left( \frac{V_s}{V} \right) K_s + \left( \frac{V - V_s}{V} \right) K_g \quad (23)$$

where  $K$  is the liquid thermal conductivity,  $K_s$  is the thermal conductivity of the solid-like degrees of freedom, and  $K_g$  is the value for gas-like degrees of freedom. Since the contribution of gas-like molecule to the liquid thermal conductivity is generally small, it is adequate to use the equation for the thermal conductivity of an ideal gas (ref. 42):

$$K_g = \frac{(9\gamma - 5)}{4} \eta_g C_{vg} \quad (24)$$

where  $\gamma$  is the ratio of the specific heat at constant pressure to that at constant volume, and  $\eta_g$  and  $C_{vg}$  are the viscosity and the heat capacity at constant volume of an ideal gas, respectively. To express  $K_s$ , we use phonon theory (refs. 43 to 45):

$$K_s = \frac{C_s^2 R}{V_s (A\omega_e^4 + BT\omega_e)} \times \frac{\left( \frac{h_D \omega_e}{kT} \right)^2 e^{-h_D \omega_e / kT}}{(e^{h_D \omega_e / RT} - 1)^2} \quad (25)$$

where  $\omega_e$  is the Einstein characteristic frequency,  $A\omega_e^4$  represents the scattering by lattice imperfections; the term  $BT\omega_e^2$  includes the normal and UmKlapp processes, and  $C_s$  is the velocity of sound in the solid. For the relatively high temperature of interest, here, we have  $h_D \omega_e / kT < 1$ , so that

$$\frac{(h_D \omega_e / kT)^2 e^{-h_D \omega_e / kT}}{(e^{h_D \omega_e / RT} - 1)^2} \simeq 1$$

Substituting equations (23) and (24), we obtain

TABLE 4.—*Calculated and Observed Sound Velocity in Liquid Argon*<sup>a</sup>

Temp, °K	$C_{calc}$ , m/sec	$C_{obs}$ , m/sec
83.85	920.5	.....
85.50	897.4	.....
85.90	.....	849.3
87.10	876.5	841.5
90.30	849.7	820.2

<sup>a</sup> Choi, D.; and Jhon, M. S.: *J. Chem. Phys.* (to be published).

$$K = \frac{RC_s^2}{V(A\omega_e^4 + BT\omega_e^2)} + \frac{V - V_s}{V} \frac{(9\gamma - 5)}{4} \eta_0 C_{vg} \quad (26)$$

The first term on the right-hand side of equation (26) represents the contribution to thermal conduction due to the vibration of molecules near their equilibrium position; and the second term, the contribution due to the random motion of molecules. Furthermore, it is convenient to absorb  $C_s$  and  $R$  into the constants  $A$  and  $B$ .

Some of the calculated results are available from reference 46.

#### Surface Tension

According to our model, there are two methods for obtaining the surface tension. One is an iteration method (ref. 47) which calculates the sum of the contribution of successive layers, and the other is a mono-layer approximation (refs. 48 and 49) which yields a simplified calculation. The former is more faithful to physical reality but requires more experimental data for calculation.

Here we describe the latter method because this procedure requires only experimental data easily available for most lubricants. According to this method, it is assumed that the surface of a liquid consists of a mono-molecular layer whose molecules have energies and entropies different from the bulk liquid. Then the partition function of the  $N_1$  surface molecules and  $N_2$  bulk molecule in the liquid is given by

$$f^{N_1+N_2} = f'^{N_1} f_B^{N_2} \quad (27)$$

where  $f'$  is the partition function for the surface molecules, and  $f_B$  is the value for the bulk liquid.

Using the equation of the surface tension  $\gamma = \left( \frac{\partial A}{\partial \Omega} \right)_{N,V,T}$  (where  $\Omega$  is

the surface area) as well as equations (13) and (27), and making the appropriate approximations (refs. 48 and 49),  $\gamma$  can be written as

$$\gamma = \omega^{-1} \left( \frac{V_s}{V} \right)^2 kT \left[ \frac{E_s - E_s'}{RT} + 3 \ln \frac{1 - \exp(-\theta'/T)}{1 - \exp(-\theta/T)} \right] \quad (28)$$

for simple liquids such as argon. Here

$$E_s' \doteq \frac{9}{12} E_s, \quad \theta' \doteq \sqrt{\frac{9}{12}} \theta, \quad \text{and} \quad \omega = \frac{\sqrt{3}}{2} \left( \frac{V_s \sqrt{2}}{N} \right)^{2/3}$$

This equation (ref. 49) is successfully applied to simple liquids, polar liquids, and molten metals at various temperatures.

#### NEEDED RESEARCH

If properties such as vapor pressure, density, and entropy are known, and if the necessary information on molecular structure is available, then the thermodynamic properties of a lubricant such as an organo-silicon, etc., can be evaluated according to the scheme discussed in the previous section. With the use of the above evaluated properties and the formula in the previous section, the sound velocity is easily evaluated. Discontinuities in the evaluated or experimental values of sound velocity may arise if there are structure changes of the lubricants.

By using the equation mentioned above, the thermal conductivity as well as the surface tension of the lubricants may be evaluated if the small amount of necessary data is available. Finally we will make a few remarks on flow properties of the lubricants. If the lubricant has Newtonian flow properties, we can calculate the viscosity of pure liquid and mixture, and the viscosity under high pressure, using the methods discussed in the previous section. It was found experimentally, however, that lubricants under pressure become thixotropic (ref. 50).

This fact indicates that molecular association is induced under high pressure and a Newtonian unit transforms into a non-Newtonian one by associating with neighbors. Since  $\Delta H$  found for the transition (ref. 51) is about 10 kcals, the reaction may involve hydrogen bond formation. The mole fraction of non-Newtonian units  $X_2$  at the transition equilibrium under pressure  $p$  is given by

$$\frac{1 - X_2}{X_2} = \exp \left( \frac{\Delta G}{RT} \right) = K_o e^{-p\Delta V/RT} \quad (29)$$

Here  $\Delta G$  is the free energy change for the transition,  $K_o$  is the equilibrium constant at  $p=0$  and  $\Delta V$  is the volume change accompanying the transition. If it is assumed that the areas occupied by the flow units are equal, then  $\eta_s$  in equation (15) can be replaced by

$$\eta_s = (1 - X_2) \eta_{\text{Newtonian}} + X_2 \eta_{\text{Non-Newtonian}} \quad (30)$$



#### ACKNOWLEDGMENT

One of the authors (M.S. Jhon) expresses his appreciation to the donors of the Petroleum Research Funds, administered by the American Chemical Society, for partial support of this research. The authors also express their appreciation to the National Science Foundation for its support.

#### DISCUSSIONS

**H. E. Sliney (NASA Lewis Research Center, Cleveland, Ohio)**

Currently there is considerable interest in the development of techniques for predicting and for measuring lubricant properties under the extreme conditions existing in concentrated contacts. These conditions include high pressure, short residence times, and high shear rates. It is well known that EHD analyses, such as those of Grubin and of Dowson and Higginson, have been a real breakthrough in this problem area, particularly with regard to predicting film thicknesses and pressure distributions throughout the contact. However, the calculations require a knowledge of lubricant viscometric properties under EHD condition and in many cases these properties are only poorly understood. The EHD approach considers the lubricating fluid as a continuum, but the conditions in concentrated contacts are such that the molecular structure of the fluid must ultimately have an effect. Therefore, it might be considered profitable to extend studies of EHD phenomena to include an atomistic approach.

The significant liquid structure theory is a step in this direction. The model used in developing the theory assumes the liquid state to be a state of matter containing both solid-like and gas-like characteristics in terms of the degrees of freedom of the molecules constituting the liquid. So-called "fluid vacancies" exist in the liquid that are analogous to, but that have more mobility than, "locked-in" vacancies in crystalline solids. Whether or not this model realistically represents the actual state of matter in a liquid, the resulting theory has apparently been successfully applied to calculating thermodynamic properties of fused salts, pure hydrocarbons, and other liquids. The theory also provides the necessary equations to calculate thermal conductivity, surface tension, and viscometric properties.

The difficulty with using the theory for calculations of interest in lubrication is that the necessary input for these calculations includes properties which are not readily available. For instance, to calculate viscosity at a given temperature and pressure, we must know vapor pressure, density, and entropy as well as details of molecular structure from which the Einstein temperature can be determined. Molar volumes must be known for the temperature and pressure under consideration. For other values of pressure and temperature, molar volumes must be

corrected by using appropriate coefficients of expansion and of compressibility.

The significant liquid structure theory is of great interest in improving understanding of the liquid state. It will be of value to the lubrication specialist for calculating viscometric properties to the degree that a reliable compilation of other properties required as input for the calculations becomes available.

**J. K. Appeldoorn (Esso Research and Engineering Company, Linden, New Jersey)**

The Eyring theory of liquid viscosity postulates that viscosity (or more precisely its reciprocal—fluidity) is determined by two factors: (1) the number of “holes” present in the liquid, and (2) the ease with which the molecules can move into these holes. The activation energy of flow can be divided into two parts: the energy required to form the hole, and the kinetic energy required for a molecule to move into the hole. Eyring (ref. 52) pointed out that it is possible to adjust temperature and pressure so that the number of holes is constant, and in this way the second (kinetic) part of the activation energy may be determined. He noted that the viscosity of many substances was in fact nearly constant at constant volume, and that 80 to 90 percent of the activation energy is therefore required to form holes, rather than to move molecules into the hole.

This agrees with the theories of Batschinski (ref. 27), Doolittle (ref. 28), and others who have proposed that the change of viscosity with temperature is due *solely* to the decrease in density, i.e., to the increase in the number of holes available.

However, for hydrocarbon molecules in the lubricating oil range, this relationship no longer holds. Viscosity is not constant at constant volume, even as a first approximation. The viscosity and density data from the ASME Pressure-Viscosity Report (ref. 50) allow an easy comparison. One such curve is given in figure 3 for a pure  $C_{25}$  hydrocarbon. The heavy lines are constant-density curves, and it is obvious that they are far from being horizontal constant-viscosity lines. At atmospheric pressure, an increase in temperature from 32° to 400° F causes a decrease in viscosity of 300-fold; a subsequent increase in pressure to give the original density only increases viscosity by 8-fold, leaving a 40-fold change still to be accounted for. This compound is not atypical. Some petroleum fractions show a 10 000-fold difference in viscosity at constant density.

Obviously one of two things is wrong: (1) constant density is not the same as constant free volume (i.e., constant holes), or (2) the kinetic energy term is far more important for large molecules than it is for simple molecules.

To examine the first of these involves estimating the volume occupied by the molecules,  $V_j$ . Free volume is then  $(V - V_j)$ . Many investigators

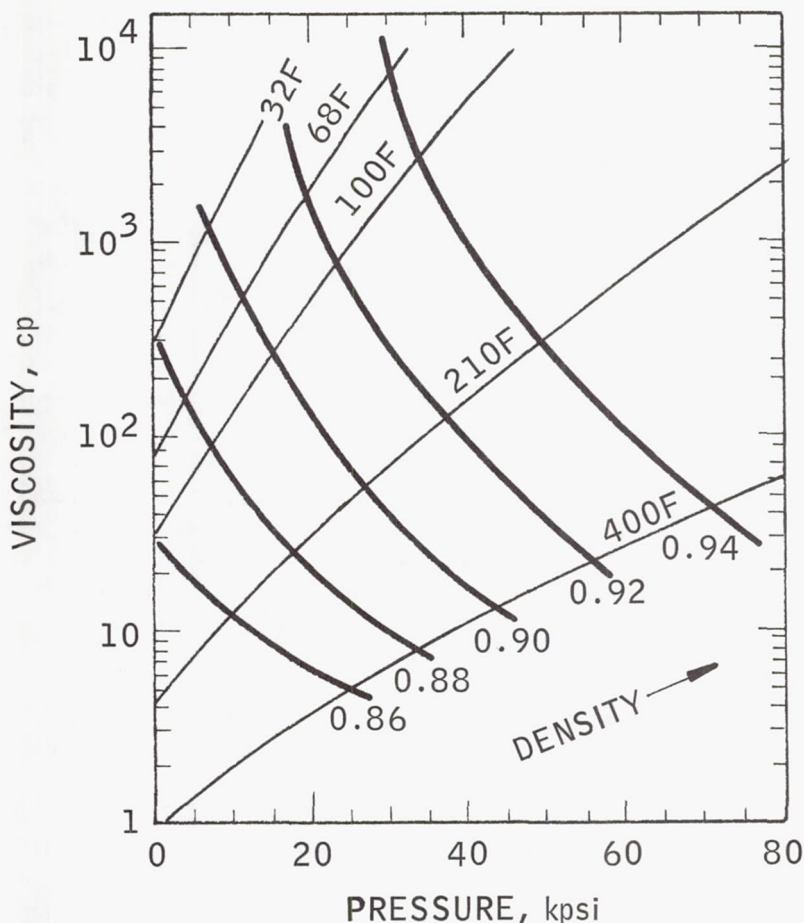


FIGURE 3.—Viscosity-density-temperature-pressure curves for 1-cyclohexyl-3 (2-cyclohexylethyl) hendecane. Note that lines of constant density are not lines of constant viscosity, even as a first approximation.

have used for  $V_j$  the van der Waals' volume,  $V_w$ , i.e., the actual volume dimensions of a molecule.  $V_w$  is not changed appreciably by either temperature or pressure, so any change in density is a direct measure of change in free volume. Figure 3 clearly shows that the use of  $V_w$  is untenable.

Other investigators have suggested using  $V_0$ , the volume of a solid crystal at  $0^\circ \text{K}$ . In such a crystal the molecules are packed together as closely as is geometrically possible, but there is some excess (although inaccessible) volume between the molecules. (For example, for hexagonally close-packed spheres, the spheres occupy 74 percent of the



volume.) Again  $V_0$  would not be expected to decrease with pressure, for this would involve the actual deformation of the molecules.  $V_0$  is therefore a constant and its use is as untenable as  $V_w$ .

Doolittle (ref. 28) proposed using an extrapolated liquid volume at  $0^\circ$  K. By plotting density vs temperature and extrapolating to  $0^\circ$  K, he obtained a volume equivalent to the volume of a liquid (having undergone no change of state) at  $0^\circ$  K. Interestingly enough, his values  $V_D$  are very close to the values calculated for  $V_0$ . At higher pressures the values of  $V_D$  are much smaller than  $V_0$ ; i.e., the molecules are packed together more closely than in a crystal at  $0^\circ$  K. This seems physically impossible, and  $V_D$  is reduced to an arbitrary adjustable constant, having no relation to a physical volume.

The significant structure theory and also that of Cohen and Turnbull (ref. 30) consider a liquid to be essentially a solid with occasional holes. The occupied volume is then the volume a crystalline or glassy solid would have, assuming it could remain a solid at that temperature and pressure. To be useful for making calculations, this requires knowing the compressibility,  $\beta$ , and thermal expansion,  $\alpha$ , of a lubricant in its solid state, and some way of extrapolating this to higher temperatures. Except for very simple compounds, these values are not known. Furthermore, in order to explain figure 3 on the basis of free volume, the ratio  $\beta/\alpha$  would have to be quite different in the solid state from that in the liquid state. Specifically the solid state would have to show much less compressibility relative to its thermal expansion. For it will be seen from figure 3 that if the oil is expanded by heating from  $32^\circ$  to  $210^\circ$  F ( $0^\circ$  to  $100^\circ$  C), it will take a pressure of only 18 000 psi to restore the original density, but 50 000 psi to restore the original viscosity (free volume).

Information on the compressibility of organic solids is not readily available. However, Bondi (ref. 53) presents data indicating that in the solid state both  $\alpha$  and  $\beta$  are  $\frac{1}{3}$  to  $\frac{1}{2}$  the values in the liquid state; i.e.,  $(\beta/\alpha)_s = (\beta/\alpha)_l$ . It can only be concluded that constant free volume does not give constant viscosity. Hogenboom, Webb, and Dixon (ref. 54) have similarly criticized the significant structure theory when applied at elevated pressures.

This leaves the conclusion that the factor controlling the viscosity of large molecules is the kinetic energy of the molecules. It appears that all theories based on free volume should be reexamined.

**V. N. Constantinescu (Mechanical Technology Incorporated, Latham, New York)**

The theory presented in the paper refers mainly to steady conditions, at least with respect to the state variables characterizing the fluid. In other words, it is considered that the fluid is in a steady state characterized by certain values for the state variables such as temperature and pressure. Under these conditions, the goal of the theory is to evaluate the transport

properties of the fluid, particularly the viscosity, as a function of the state variables. It is also assumed that the viscosity will vary with pressure and temperature, for example, in accordance with equation (19), i.e., viscosity follows instantaneously the changes in the state variables.

The following question seems important: what are the conditions for the validity of this quasi-steady treatment when it is applied to phenomena occurring in a rather short time (of the order of  $10^{-6}$  to  $10^{-8}$  sec)?

Indeed, when we want to study the changes occurring in a fluid passing through an elastohydrodynamic contact, we may recognize that an element of the fluid is subjected to an increase in shear stress, pressure, and temperature with a duration which is of the order of  $t_0 \sim \ell/U$  where  $\ell$  is the length of the contact Hertzian ellipse and  $U$  is the rolling speed. As  $U$  is of the order of magnitude of  $10^2$  to  $10^3$  in/sec and  $\ell \sim 10^{-3}$  in., it is seen that  $t_0 \sim 10^{-6}$  sec or even less. Indeed, within the interval  $t_0$ , the pressure, for example, raises from 10 to 20 psi to 200 000 to 400 000 psi and then falls again to the atmospheric value.

In order to make the question more relevant, suppose we are able to perform the following experiment: we suddenly raise the pressure from  $p_1$  to  $p_2$  as shown in figure 4, and we are able to measure instantaneously the viscosity. As in any system, it is reasonable to assume that the dynamic response cannot rigorously be a step function, but rather the viscosity will reach the new level  $\eta_2$  after a certain time  $T_0$ . Within the interval  $T_0$  it is probable that the viscosity will gradually increase from  $\eta_1$  to  $\eta_2$  (and eventually may exhibit some oscillatory behavior before stabilizing to the new level  $\eta_2$ ).

What is the duration of this time interval  $T_0$  for complex fluids like oils? If the speed of sound  $c$  is a measure of how quickly the transport properties may be changed, then as  $c \sim 10^5$  in/sec and  $\ell \sim 10^{-3}$  in, perhaps  $T_0 \sim 10^{-8}$  sec.  $T_0$  may be much smaller for simple fluids like liquid argon, but may be greater for complex liquids as oils or polymers. Indeed, especially when a sudden, big rise in pressure level is occurring, certain changes in the structure of the liquid, leading to changes in viscosity, may be expected. But all these changes will require a certain time  $T_0$ , which probably will be a function of such parameters as temperature, pressure  $p_1$ , and pressure increment  $\Delta p$ .

If  $T_0$  and  $t_0$  are of the same order of magnitude, then the discrepancies between theory and experiment existing in the elastohydrodynamic lubrication may well be attributed to the delay in viscosity response. Indeed, a chart giving viscosity  $\eta$  versus pressure  $p$  may then have the aspect presented qualitatively in figure 5. Thus, where the time allowed to obtain the pressure increment is large, the usual  $\eta(p)$  curve is obtained ( $t_0 \rightarrow \infty$ ); however, for short time interval  $t_0$ , the viscosity pressure dependence will continuously depart from the previous static curve. In the limit when  $t_0 \rightarrow 0$ , it is probable that the viscosity will exhibit a negligible

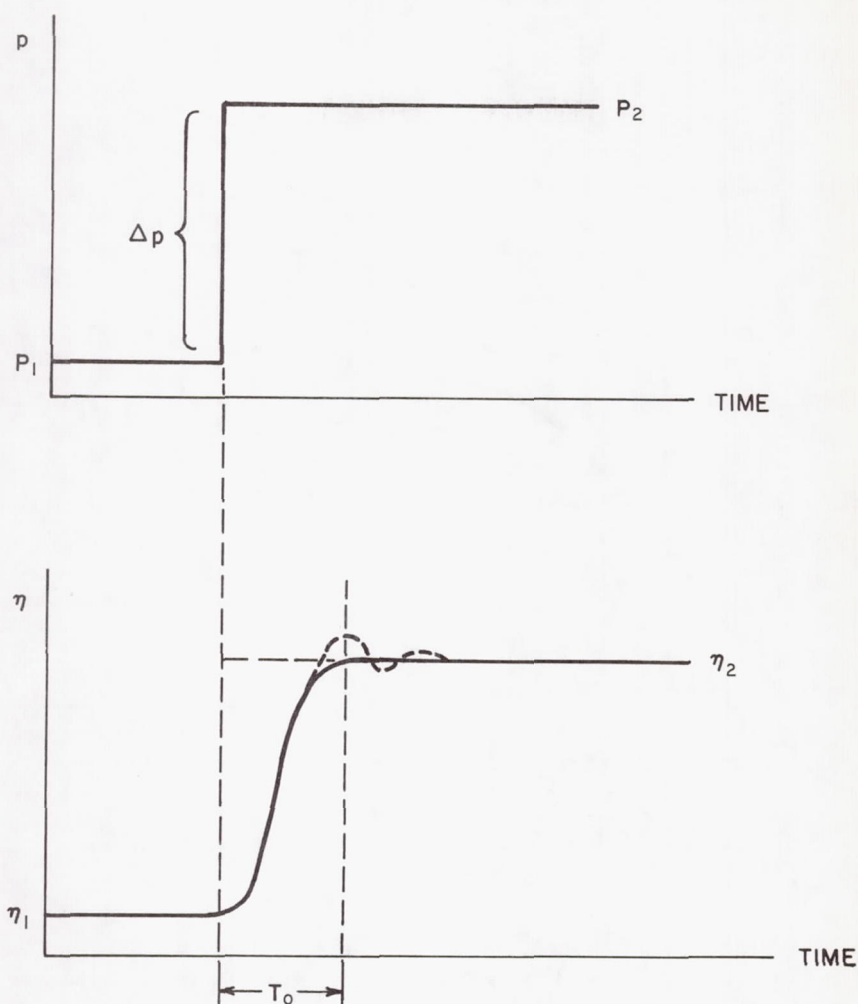


FIGURE 4.—Possible response in viscosity variation to a sudden pressure increment.

change (the liquid has not time to “see” the change in pressure). Such a dependence may be empirically expressed (at constant temperature) as

$$\eta = \eta_0 e^{\alpha p} \quad \text{for } p < p'(t_0)$$

$$\eta = \eta' e^{\alpha'(p-p')}; \quad \alpha' = \alpha'(t_0)$$

which agrees with some existing experimental evidence concerning the effective mean viscosity in a lubricated contact.



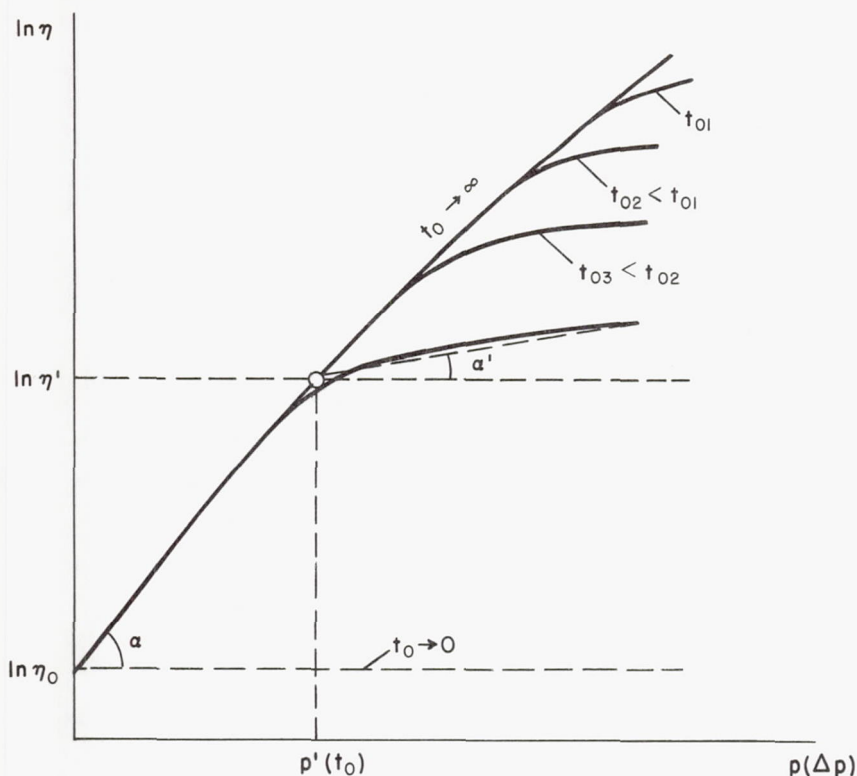


FIGURE 5.—Possible dependence of viscosity versus pressure.

## LECTURERS' CLOSURE

As Mr. Sliney mentioned, there seems to be no liquid which cannot be usefully examined using the significant liquid structure model. However, the few necessary input data must be known to evaluate the thermodynamic and transport properties of lubricants or other liquids. Unfortunately, sufficient data for the lubricants except for simple hydrocarbons such as  $C_5$ – $C_{10}$ ,  $C_{16}$ , and  $C_{17}$ , etc, are not readily available at the present time.

Our model will be of great value to the lubrication scientists when the necessary input data become available, and even at present should be very useful in singling out the factors which influence viscosity.

Eyring's early model has a limited range of applicability as Mr. Appeldoorn indicated. However, according to the newer viscosity theory, developed by Eyring et al. in terms of the significant structure liquid model,

the equation for shear viscosity has the form  $\gamma = (\nu_s/\nu)\gamma_s + \frac{(\nu - \nu_s)}{\nu}\gamma_0$ ,

where  $\nu_s$  is the specific volume of the solid and  $\nu$  is the specific volume of the liquid. The quantity  $\gamma_\theta$  is a gas-like term derived from kinetic theory

in the form  $\gamma_\theta = (5/16d^2) \left( \frac{mkt}{\pi} \right)^{1/2}$ , where  $m$  and  $d$  are the molecular mass

and diameter, respectively.

Since  $\gamma_\theta$  is small, this contribution can be neglected. In this case the viscosity expression (ref. 55) is

$$\gamma \simeq \gamma_s = \frac{\nu}{\nu_s} \frac{(\pi m k T)^{1/2} N \ell_f}{2(\nu - \nu_s) K} \exp \left[ \frac{-a' V_s Z \phi(a)}{(V - V_s) 2kT} \right] \quad (31)$$

where  $\ell_f$  is a constant representing the free distance between nearest neighbors,  $K$  is a constant transmission coefficient,  $a'$  is an adjustable constant,  $\phi(a)$  is an intermolecular potential function,  $Z$  is a constant equal to the number of nearest neighbors in the lattice,  $k$  is Boltzmann's constant, and  $N$  is Avogadro's number. Equation (31) may be rewritten as

$$\ell_f [\gamma (V - V_s) / T^{1/2}] = A + [B \nu_s / (\nu - \nu_s) T] \quad (32)$$

Here  $A$  and  $B$  are constant at atmospheric pressure.

As one sees,  $\gamma$  is a function of  $T$  and  $V$ . Equation (31) or (32) is quite satisfactory for many liquids (ref. 56).

A satisfactory model should explain the pressure dependence of the viscosity as Hogenboom (ref. 54) and Appeldoorn mentioned. Most recently the explicit functional expressions of the molar solid-like volume,  $V_s$ , and the intermolecular potential,  $\phi(a)$ , in terms of the pressure,\* were introduced in equation (31). The results quite satisfactorily represent the viscosity behavior of the liquid hydrocarbons such as n-C<sub>12</sub>, n-C<sub>15</sub>, and n-C<sub>18</sub>, etc. The following is a more detailed discussion of this effect.

The effect of pressure on the solid-like volume  $\nu_s$  can be considered as  $V_s = V_{s0}(1 - \beta \Delta p)$ , where  $\beta$  is the compressibility of solid-like molecules and  $\Delta p$  is the pressure difference between 1 atm and the applied pressure,  $p$ . This means that the size of holes are not constant, but are a function of the pressure.

Since  $\ell_f$  is equal to  $2 \left[ \frac{(\sqrt{2} V_s)^{1/3}}{N} - a \right]$ , where the collision diameter  $\alpha$  is

related to  $\alpha \simeq (b/4N)^{1/3} = (V_c/12N)^{1/3}$ , where  $b$  is Van der Waal's constant, where  $V_s$  is the solid like molar volume and  $V_c$  is the critical molar volume,  $A$  is proportional to  $\ln \ell_f$ . For the calculation of  $B$ , the Lennard-Jones (12-6) potential is used for the expression of  $\phi(a)$ . Then,

\* Jhon, M. S.; Klatz, W. L.; and Eyring, H.: J. Chem. Phys. (to be published).

$$\begin{aligned}\phi(a) &= \epsilon [1.0109 (Na^3/V_s)^4 - 2.4090 (Na^3/V_s)^2] \\ &= \epsilon [1.0109 (V_c/12V_s)^4 - 2.4090 (V_c/12V_s)]\end{aligned}$$

In addition, appropriate consideration of the fraction of non-Newtonian flow may improve the results.

Now, viscosity can be, in general, replaced by

$$\gamma = (1 - X_2)\gamma_{\text{Newtonian}} + X_2\gamma_{\text{Non-Newtonian}}$$

Here  $X_2$  is the mole fraction of non-Newtonian units (see our text).

The questions raised by Dr. Constantinescu have the following two points. Is the viscosity equation still available for phenomena occurring in a rather short time such as the flow at the elastohydrodynamic contact? What is the limited time duration for complex fluids like oil?

Our viscosity equation is based on the validity of the quasi-steady treatment. Hence, it is meaningless to calculate the viscosity for shorter times than the relaxation time of viscous flow.

Within the limited time interval  $t$ , the viscosity will gradually increase from  $y_1$  to  $y_2$ . The values of  $y_1$  and  $y_2$  can be evaluated theoretically. For complex liquids such as oils or polymers,  $t$  may be much greater than for the simple liquids. The calculated relaxation time is  $\beta = 9 \times 10^{-3}$  sec (ref. 56) for high polymers; this is a reasonable value for this time of duration.

#### NOMENCLATURE

$f_N$	partition function; a mathematical equation which describes the distribution of molecules in different energy states.
$\psi(r)$	Lennard-Jones (6-12) potential
$T$	temperature °K
$R$	gas constant
$k$	Boltzmann's constant
$m$	mass of a particle
$h$	Planck's constant
$\Delta S_f$	entropy of fusion
$L$	number of cells
$N$	number of particles
$V_s$	molar volume of the solid at the melting point
$V$	molar volume of the liquid
$C_v$	molar specific heat at constant volume
$C_p$	molar specific heat at constant pressure
$E_s$	heat of sublimation
$\theta$	Einstein characteristic temperature
$V_m$	liquid molar volume at the melting point
$n$	number of nearest neighbors around a molecule at the melting point; equal to $12 V_s/V_m$



$a', b$ and $a$	dimensionless constant
$P$	pressure
$A$	Helmholtz free energy
$S$	entropy
$E$	internal energy
$\eta$	viscosity
$\Delta G^\ddagger$	free energy of activation
$\Delta H^\ddagger$	enthalpy of activation
$\Delta S^\ddagger$	entropy of activation
$v_f$	free volume
$f'$	shear stress
$\dot{S}$	rate of strain
$K'$	transmission coefficient
$D$	diffusion coefficient
$C$	Sound velocity
$\gamma'$	$\frac{C_p}{C_v}$
$\beta_T$	isothermal compressibility
$\rho$	density
$M$	molecular weight
$\alpha$	expansion coefficient
$K$	thermal conductivity
$\omega_e$	Einstein characteristic frequency; equal to $2\pi\nu$
$\gamma$	surface tension
$\sigma$	collision diameter
$\epsilon$	energy parameter
$h_D$	$\frac{h}{2\pi}$
$\nu$	the frequency
$\Omega$	surface area
$X_s$	mole fraction of solid-like molecule
$X_g$	mole fraction of gas
$d$	diameter of the molecule
$r$	$ZK'$
$Z$	coordination number of the neighboring position around molecule
$\xi$	effective number of neighbors of a molecule lying in the same plane.
$K_0$	the equilibrium constant at $P=0$
Suffix 1 or 2	denotes the species 1 and 2, respectively
Suffix $s$ or $g$	denotes the solid-like molecule, and the gas-like molecule, respectively

## REFERENCES

1. EYRING, H.: Viscosity, Plasticity and Diffusion as Examples of Absolute Reaction Rates. *J. Chem. Phys.*, vol. 4, 1936, p. 283.
2. BARKER, J. A.: *Lattice Theories of the Liquid State*. The MacMillan Company, 1963.
3. PRIGOGINE, I.: *The Molecular Theory of Solutions*. Interscience Publishers, 1957, pp. 53, 99, 115, 374.
4. EYRING, H.; REE, T.; AND HIRAI, N.: Significant Structures in the Liquid State, I. *Proc. Natl. Acad. Sci.*, vol. 44, 1958, p. 683.
5. EYRING, H.; AND JHON, M. S.: *Significant Liquid Structures*. John Wiley and Sons, Inc., 1969.
6. MAYER, J. E.; AND MAYER, M. G.: *Statistical Mechanics*. John Wiley and Sons, Inc., 1950, p. 295.
7. KIRKWOOD, J. G.; LEWINSON, V. A.; AND ALDER, B. J.: Radial Distribution Functions and the Equations of State of Fluids Composed of Molecules Interacting According to Lennard-Jones Potential. *J. Chem. Phys.*, vol. 20, 1952, p. 929.
8. FRISCH, H. L.: *Advances in Chemical Physics*. Vol. 6. Interscience Publishers, 1964, p. 229.
9. PERCUS, J. K.; AND YEVIK, G. J.: Analysis of Classical Statistical Mechanics by Means of Collective Coordinates. *Phys. Rev.*, vol. 110, 1958, p. 1.
10. BUEHLER, R. J.; WENTORF, R. H.; HIRSCHFELDER, J. O.; AND CURTISS, C. F.: The Free Volume for Rigid Sphere Molecules. *J. Chem. Phys.*, vol. 19, 1951, p. 61.
11. EYRING, H.; AND HIRSCHFELDER, J. O.: The Theory of the Liquid State. *J. Phys. Chem.*, vol. 41, 1937, p. 249.
12. LENNARD-JONES, J. E.; AND DEVONSHIRE, A. F.: Critical Phenomena in Gases. I. *Proc. Roy. Soc. (London)*, vol. 163, 1937, p. 53.
13. LENNARD-JONES, J. E.; AND DEVONSHIRE, A. F.: Critical Phenomena in Gases. II. Vapor Pressures and Boiling Point. *Proc. Roy. Soc. (London)*, vol. 165, 1938, p. 1.
14. HIRSCHFELDER, J. O.; STEVENSON, D.; AND EYRING, H.: Theory of Liquid States. *J. Chem. Phys.*, vol. 5, 1937, p. 896.
15. CERNUSCHI, F.; AND EYRING, H.: An Elementary Theory of Condensation. *J. Chem. Phys.*, vol. 7, 1939, p. 547.
16. ROWLINSON, J. S.; AND CURTISS, C. F.: Lattice Theories of the Liquid States. *J. Chem. Phys.*, vol. 19, 1951, p. 1519.
17. ONO, S.: *Mem. Fac. Eng. Kyushu Univ. (Japan)*, vol. 16, 1947, p. 190.
18. PEEK, H. M.; AND HILL, T. L.: On Lattice Theories of the Liquid States. *J. Chem. Phys.*, vol. 18, 1950, p. 1252.
19. HENDERSON, D.: Hole Theories of Liquids and Dense Gases. I. Equation of State. *J. Chem. Phys.*, vol. 37, 1962, p. 631.
20. BARKER, J. A.: The Tunnel Theory of Fluids, the 12-6 Fluids. *Proc. Roy. Soc. (London)*, vol. A259, 1961, p. 442.
21. EISENSTEIN, A.; AND GINGRICH, N.: Diffraction of X-rays by Argon in the Liquid, Vapor, and Critical Regions. *Phys. Rev.*, vol. 62, 1942, p. 261.
22. WALTER, J.; AND EYRING, H.: A Partition Function for Normal Liquids. *J. Chem. Phys.*, vol. 9, 1941, p. 393.
23. GROSH, J.; JHON, M. S.; REE, T.; AND EYRING, H.: On An Improved Partition Functions of Significant Structure Theory. *Proc. Natl. Acad. Sci.*, vol. 57, 1967, p. 1566.

24. PIEROTTI, R. A.: On the Significant Structure Theory of Liquids. *J. Chem. Phys.*, vol. 43, 1965, p. 1072.
25. STEWART, C. W.; AND VON FRANKENBERG, C. A.: An Application of Significant Structures Theory to Amorphous Polyethylene. *J. Polymer. Sci.*, vol. 5, 1967, p. 623.
26. STEWART, C. W.; AND VON FRANKENBERG, C. A.: Significant Structure Theory of the Surface Tension of Polyethylene. *J. Polymer. Sci.*, vol. 6, 1968, p. 1686.
27. BATSCHINSKI, A. J.: Untersuchungen über die innere Reibung der Flüssigkeiten. *I. Z. Physik. Chem.*, vol. 84, 1913, p. 643.
28. DOOLITTLE, A. K.: Studies in Newtonian Flow. II. The Dependence of the Viscosity of Liquids on Free Space. *J. Appl. Phys.*, vol. 22, 1951, p. 1471.
29. DOOLITTLE, A. K.: Studies in Newtonian Flow. III. The Dependence of the Viscosity of Liquids on Molecular Weight and Free Space. *J. Appl. Phys.*, vol. 23, 1952, p. 238.
30. COHEN, M. H.; AND TURNBULL, D.: Molecular Transport in Liquids and Glasses. *J. Chem. Phys.*, vol. 31, 1959, p. 1164.
31. BEUCHE, F.: Mobility of Molecules in Liquids Near the Glass Temperature. *J. Chem. Phys.*, vol. 30, 1959, p. 748.
32. MCCEDO, P.; AND LITOVITZ, T. A.: On the Relative Roles of Free Volume and Activation Energy in the Viscosity of Liquids. *J. Chem. Phys.*, vol. 42, 1965, p. 245.
33. WILLIAMS, M. L.; LANDEL, R. F.; AND FERRY, J. D.: The Temperature Dependence of Relaxation Mechanism in Amorphous Polymers and Other Glass Forming Liquids. *J. Am. Chem. Soc.*, vol. 77, 1955, p. 3701.
34. MOOR, W. J.: *Physical Chemistry*. Prentice-Hall, Inc., p. 118.
35. REE, T. S.; REE, T.; AND EYRING, H.: Significant Liquid Structure Theory. IX. Properties of Dense Gases and Liquids. *Proc. Natl. Acad. Sci.*, vol. 48, 1962, p. 501.
36. EYRING, H.; REE, T.; AND HIRAI, N.: The Viscosity of High Polymers—The Random Walk of a Group of Connected Segments. *Proc. Natl. Acad. Sci.*, vol. 44, 1958, p. 1213.
37. HIRAI, N.; AND EYRING, H.: Bulk Viscosity of Liquid. *J. Appl. Phys.*, vol. 29, 1958, p. 810.
38. CHANG, S.; PAK, H.; PAIK, W. K.; PAE, S.; JHON, M. S.; AND AHN, W. S.: Modified Theory of Liquid Structure (Seoul Natl. Univ., Korea). *J. Korean Chem. Soc.*, vol. 8, 1964, pp. 31-33.
39. PAIK, W. K.; AND CHANG, S.: Structure and Properties of Liquid Carbon Tetrachloride. *J. Korean Chem. Soc.*, vol. 8, 1964, pp. 30-32.
40. TIMMERMAN, J.: *Physicochemical constants of Pure Organic Compounds*. Vol. 1, Elsevier, 1963.
41. HERIC, E. L.; AND BREWER, J. G.: Viscosity of Some Binary Liquid Nonelectrolyte Mixtures. *J. Chem. Eng. Data*, vol. 12, no. 4, 1967, pp. 574-583.
42. LOEB, L. B.: *Kinetic Theory of Gases*. McGraw-Hill Book Co., Inc., ch. 6, 1927.
43. KLEMENS, P. G.: Thermal Conductivity of Dielectric Solids at Low Temperatures. *Proc. Roy. Soc. (London)*, vol. A208, 1951, p. 108.
44. CALLAWAY, J.: Model for Lattice Thermal Conductivity at Low Temperatures. *Phys. Rev.*, vol. 113, 1959, p. 1046.
45. PEIERLS, R.: Zur Kinetischen Theorie der Wärmeleitung in Kristallen. *Ann. Phys.*, vol. 3, 1929, p. 1055.
46. LIN, S. H.; EYRING, H.; AND DAVIS, W. J.: Thermal Conductivity of Liquids. *J. Phys. Chem.*, vol. 68, 1964, p. 3017.
47. CHANG, S.; REE, T.; EYRING, H.; AND MATZNER, I.: Progress in International



- Research on Thermodynamics and Transport Properties. Masi and Tsai, eds., Academic Press, 1962, p. 88.
48. REE, T. S.; REE, T.; AND EYRING, H.: Significant Structure Theory of Surface Tension. *J. Chem. Phys.*, vol. 41, 1964, p. 524.
  49. LU, W. C.; JHON, M. S.; REE, T.; AND EYRING, H.: The Significant Structure Theory Applied to Surface Tension. *J. Chem. Phys.*, vol. 46, 1967, p. 1075.
  50. HERSEY, M. D.; AND HOPKINS, R. T.: Viscosity of Lubricants Under Pressure. ASME (New York), 1954.
  51. HAHN, S. J.; EYRING, H.; HIGUCHI, L.; AND REE, T.: Flow Properties of Lubricating Oils under Pressure. *NLGI Spokesman*, June 1958, p. 121.
  52. GLASSTONE, S.; LAIDLER, K. J.; AND EYRING, H.: The Theory of Rate Processes. McGraw-Hill Book Co., Inc., 1941, p. 486.
  53. BONDI, A.: Physical Properties of Molecular Crystals, Liquids, and Glasses. John Wiley and Sons, Inc., 1968.
  54. HOGENBOOM, D. L.; WEBB, W.; AND DIXON, J. A.: Viscosity of Several Liquid Hydrocarbons as a Function of Temperature, Pressure, and Free Volume. *J. Chem. Phys.*, vol. 46, 1967, p. 2586.
  55. REE, T. S.; REE, T.; AND EYRING, H.: Significant Structure Theory of Transport Phenomena. *J. Phys. Chem.*, vol. 68, 1964, p. 3262.
  56. EYRING, H.; REE, T.; GRANT, D. M.; AND HIRST, R. C.: *Zeitschrift fur Electrochemie*, vol. 64, 1960, p. 146.

# The Rheological Behavior of Lubricants

H. NAYLOR

Shell Research Limited  
Thornton Research Center  
Chester, England

Elastohydrodynamic theory developed over the last 20 years has been remarkably successful in explaining the many features of the behavior of heavily loaded lubricated contacts. Problems remain, however, and several of the important ones center on the rheological properties of the lubricant at the high pressures and high rates of shear to which it is exposed in such contacts.

This paper discusses a number of rheological models that have been applied in attempts to explain observed phenomena, particularly those involving lubricant film thickness and frictional tractions, and examines the extent to which each has been successful. It is concluded that existing knowledge of the rheological properties of lubricants is, in every case, insufficient to allow critical evaluation.

**E**LASTOHYDRODYNAMIC LUBRICATION (EHL) has proved to be both the most important and the most exciting development in the lubrication field in this century: important because it is fundamental to a wide range of engineering technology, the extent of which we are only just beginning to explore; exciting because, although conceptually simple, it has turned out to involve quite unexpected mathematical and physical complexities.

Many of these complexities center on the behavior of the fluids used as lubricants at the extreme conditions of temperature, pressure, and deformation rate to which they are subjected in an elastohydrodynamically lubricated contact. Also, in attempts to provide explanations for the observed phenomena, our knowledge of the physical properties of matter is being stretched to the limit—in some cases, beyond.

In the 20 years since an acceptable description of EHL was first provided, tremendous progress has been made on both theoretical and experimental fronts. The extent of this progress has been very adequately presented at this symposium. It is the purpose of this contribution to review the available knowledge of the behavior of lubricants, particularly

with regard to their rheological properties, in relation to the lubrication of heavily loaded contacts. The subject, in the spirit of this symposium, is a truly interdisciplinary one involving mathematicians, rheologists, chemists, and engineers.

The engineering problem is, in principle, very simple. We seek a three-dimensional representation of lubricant viscosity of the form

$$\eta = \phi(P, T, D) \quad (1)$$

where  $P$  is the pressure,  $T$  the temperature, and  $D$  the shear rate, the representation to be valid for the whole range of the variables in which we are interested. We should like this representation to be capable of interpretation in molecular terms.

The effect of the first two factors,  $P$  and  $T$ , on the properties of Newtonian lubricants has been the subject of much study over many years and, although much is known, much still remains to be done. The problems will not be dealt with here.

The introduction of the third factor, the shear rate,  $D$ , greatly increases the complexity of the problem, and it is only in the last decade or so that real progress, in the present context, has been made. Much of the difficulty stems from the fact that we have very limited access to experimental data on the behavior of lubricants at very high shear rates. Indeed, it is only as a consequence of work on EHL undertaken in equipment such as the disc machine that many of the problems have come to light.

#### EXPERIMENTAL OBSERVATIONS

At the risk of some duplication, it is convenient to take as a starting point an examination of those experimental results that cannot be adequately explained in terms of Newtonian elastohydrodynamic theory. These fall under two headings: the relatively straightforward matter of measured film thickness for lubricants that display non-Newtonian behavior under laboratory test conditions and the more interesting problems presented by observed frictional behavior in all EHL situations.

##### Film Thickness

A number of workers have studied the relationship between observed and theoretical film thicknesses (refs. 1 to 3), and figure 1 shows a typical set of results for a Newtonian lubricant. The fall in observed film thickness at high rolling speeds may be attributed to viscous heating in the inlet zone of the contact not being taken into account in the calculations. This has been discussed by Cheng and Orcutt (ref. 4). In lubricants that display non-Newtonian behavior in continuous-shear laboratory viscometers, the observed film thickness falls, not unexpectedly, below that calculated on the basis of the zero shear rate viscosity. Actual measurements on such lubricants (of which high-molecular-weight-polymer-containing oils and



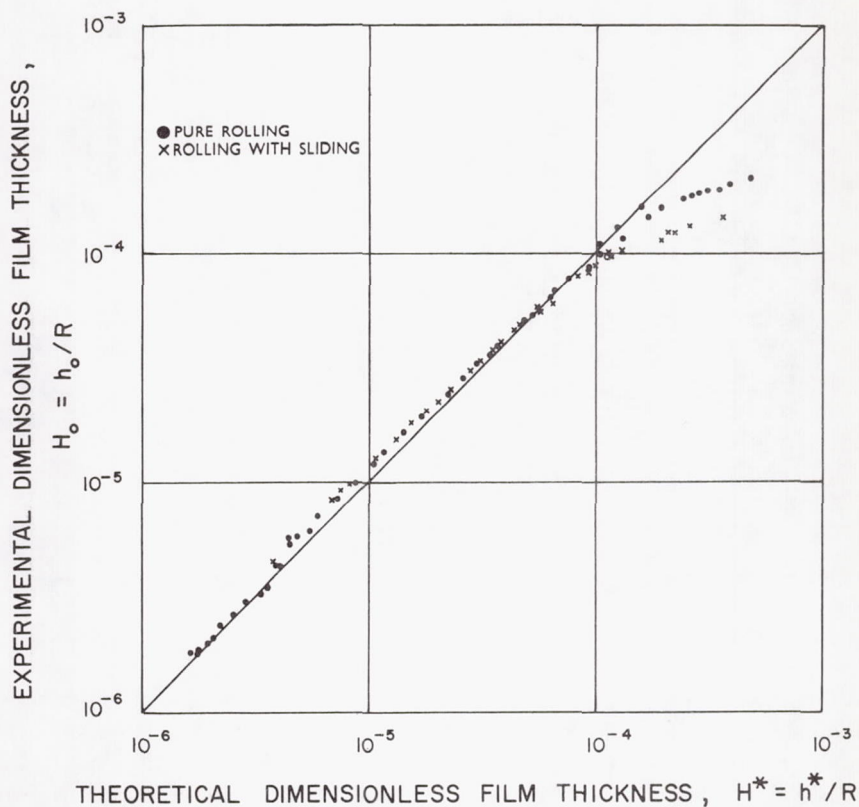


FIGURE 1.—Comparison of measured non-dimensional oil-film thickness  $H_o = h_o/R$  with predicted values  $H^* = h^*/R$  (ref. 3).

the silicones are examples) were first reported by Sibley and Orcutt (ref. 2). Further experiments have been reported by Archard and Kirk (ref. 5) and by Dyson and Wilson (ref. 6). Results obtained by the latter workers on a series of polydimethylsiloxanes are shown in figure 2.

However, no mystery attaches to these observations since, as already stated, the non-Newtonian properties of these liquids can be demonstrated in laboratory viscometers at room temperature and atmospheric pressure. Nevertheless, adequate representation of their behavior in the terms of equation (1) may still present problems.

#### Friction

An adequate theory of hydrodynamic lubrication must be capable of predicting not only the film thickness but also frictional tractions at the opposing surfaces. It is in this area that EHL theory has proved to be inadequate. The outstanding problems may be stated as follows. Film

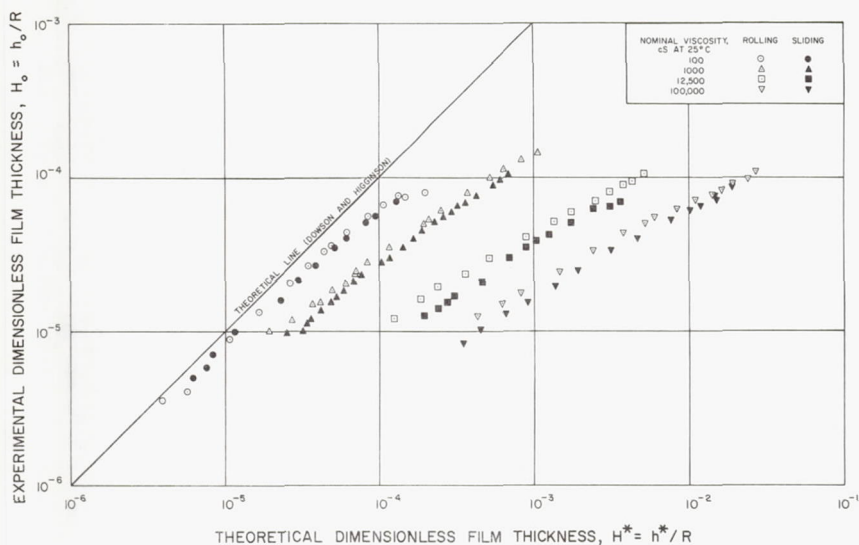


FIGURE 2.—Comparison of experimental film thicknesses given by silicone fluids with predictions based on a Newtonian theory (ref. 6).

thickness is governed by the conditions in the inlet region where pressures and shear rates are modest. The frictional tractions, on the other hand, are determined by the physical properties of the lubricant contained within the actual contact area. As Crook (ref. 7) showed we may regard the Hertzian region as forming a parallel-plate viscometer and estimate the effective viscosity from the known film thickness and the observed tractions. If the film thickness and load are maintained constant, the curve of traction as a function of sliding speed has the familiar form of figure 3. In general it is possible to define three regions as shown. At low sliding speeds a linear relation exists, the slope of which defines a quasi-Newtonian viscosity, and the behavior is isothermal. At high sliding speeds the traction decreases as sliding speed increases and this can be attributed to some extent to the influence of temperature on viscosity. In the transition region, thermal effects provide a totally inadequate explanation since the observed traction may be several orders of magnitude lower than the calculated values even when temperature effects are considered.

To this already complicated situation two further effects must be added. First, in the linear region it is found that the viscosity deduced from the slope of traction against sliding speed decreases with increasing rolling speed. Second, recent experimental results obtained over wide ranges of load and speed indicate the existence of an upper limit of frictional traction (ref. 8) as shown in figure 4, which also shows the effect of rolling speed at low sliding speeds.

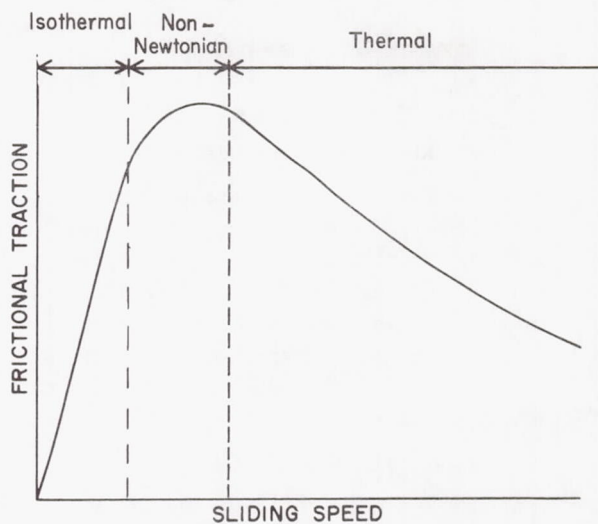
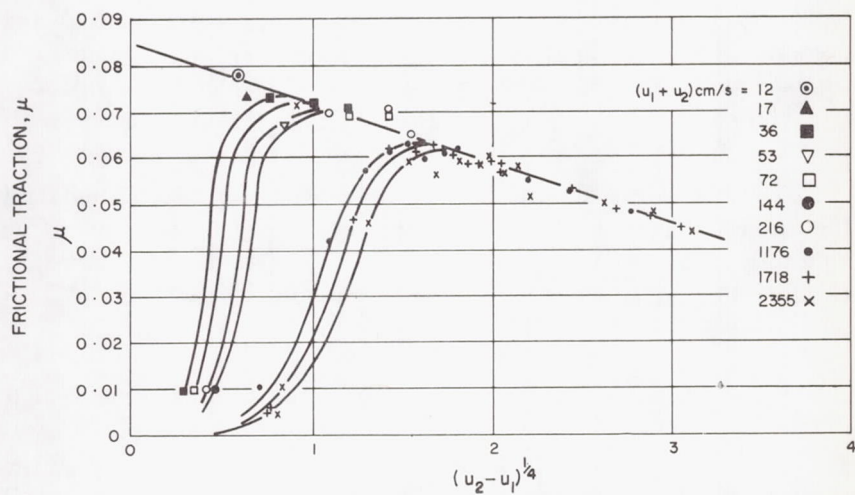


FIGURE 3.—Plot of friction against sliding speed.

ROLLER RADIUS	11.4 cm
ROLLER CROWN RADIUS	11.4 cm
ROLLER SURFACE TEMPERATURE	27° C
MEAN HERTZIAN PRESSURE	$15.47 \times 10^9 \text{ dyn/cm}^2$

FIGURE 4.—Variation of coefficient of friction with (sliding velocity)<sup>1/4</sup> (ref. 9).



Clearly these observations cannot be explained in terms of a Newtonian lubricant with viscosity a function of temperature and pressure only. Some non-Newtonian behavior must be invoked, and it is our task now to enquire into its nature.

#### RHEOLOGICAL DESCRIPTION

Attempts to explain some or all of these observed phenomena have been based on two alternative approaches: (1) the application of some rheological model of the liquid state leading to a relationship of the form of equation (1) which will predict correctly a continuously changing effective viscosity in terms of the independent variables  $P$ ,  $T$ , and  $D$ , and (2) the abandonment of the implications of equation (1) in favor of a new description of lubricant behavior, e.g., the plastic solid concept of Smith (ref. 8) and Plint (refs. 9 and 10).

#### The "Power Law" Fluid

The simplest possible approach is to take the Newtonian equation of state

$$\tau = \eta \frac{\partial u}{\partial y}$$

where  $\tau$  is shear stress,  $\eta$  the viscosity,  $u$  the velocity, and  $y$  the coordinate normal to the flow direction. Then a nonlinear dependence on shear rate is introduced empirically; for example,

$$\tau = \left( \phi \frac{\partial u}{\partial y} \right)^m$$

Such relationships can be combined with descriptions of temperature and pressure effects in a variety of ways. They have frequently been used to combine temperature and shear rate, as in the expression used by Martin (ref. 11); viz,

$$\eta = D^n \exp(-BT)$$

However, such rheological equations of state depend on measured data for the determination of the disposable material constants. Since the values of  $P$ ,  $T$ , and  $D$  of interest in EHL situations are inaccessible to laboratory viscometers, they are not usually of value except in a qualitative way.

#### The Ree-Eyring Model

The best known approach to the problem in molecular terms is that initiated by Eyring, which leads to an expression for viscosity of the form:

$$\eta = \sum_{p=1}^p \frac{x_n \beta_n}{\alpha_n} \frac{\sinh^{-1} \beta_n D}{\beta_n D} \quad (2)$$

Based on Eyring's relaxation theory of viscous flow, it seeks to derive the transport properties of a fluid directly from a description of the molecular processes involved. However, owing to the inaccessibility of the disposable parameters, it can be applied to real liquids only in the same general way as the "power-law" equations, i.e., the constants must be determined from data obtained from continuous shear experiments made at the appropriate conditions.

Equation (2) has been shown to give good agreement with the measured flow behavior of a variety of non-Newtonian fluids (ref. 12). To obtain such agreement it may be necessary to take  $p = 3$  and include an additional term representing a Newtonian fraction. When this is done the resulting rheological equation of state is too cumbersome to be useful analytically.

The Ree-Eyring model has been applied to the elastohydrodynamic system using the simplified form

$$\tau = \frac{X}{\alpha} \sinh^{-1} (\beta D) \quad (3)$$

The results will be discussed in a later section.

#### Viscoelasticity

It is to be expected that all liquids subjected to transient shear stresses of sufficiently short duration will exhibit some measure of elastic behavior. This elastic response constitutes a deformation mode additional to the more familiar viscous response. Consequently a liquid that exhibits both effects simultaneously appear "softer" than would a corresponding Newtonian liquid.

The well-known Maxwell model for such behavior assumes a stress-strain relationship of the form

$$\tau + \lambda \frac{\partial \tau}{\partial t} = \eta D$$

where  $\lambda$  is the relaxation time,  $\eta/G_\infty$ , and which for a sinusoidal shear stress may be written

$$\tau (1 + j\omega\lambda) = j\omega\eta S$$

where  $S$  is the shear strain: at low frequencies,  $\tau = j\omega\eta S$ ; and at high frequencies,  $\tau = \eta S/\lambda$ . Thus  $\eta/\lambda$  defines an elastic rigidity modulus which at limiting high frequency is

$$G_\infty = \eta_0/\lambda$$

In general, the ratio of shear stress to shear strain,  $\tau/S$ , defines a shear modulus,  $G^*$ , which is complex, i.e.,

$$G^* = G' + jG''$$

Thus

$$\frac{G'}{G_\infty} = \frac{\omega^2 \lambda^2}{1 + \omega^2 \lambda^2}, \quad \frac{G''}{G_\infty} = \frac{\omega \lambda}{1 + \omega^2 \lambda^2}$$

and the apparent viscosity

$$\eta' = \frac{G''}{\omega}$$

It has been shown (refs. 13 to 15) that for the Maxwell model, identity can be established between the behavior in oscillatory and that in continuous shear such that

$$\eta' = \frac{G''(\omega)}{\omega} = \frac{G''(D)}{D}$$

The technique of Mason (ref. 16) as developed by others—notably Barlow and Lamb (ref. 17)—enables the shear mechanical impedance of liquids to be measured over the relaxation region and hence the moduli  $G^*$ ,  $G'$  and  $G''$  to be deduced. Thus, although this approach still depends on measured properties for its application, it is the only approach available that provides access to the viscometric properties of liquids at very high shear rates.

As we shall see later, there are certain objections to the simple Maxwell model. Barlow, Erginsav, and Lamb (ref. 18) have recently proposed a new model that appears to be in much better agreement with the behavior of real liquids. In this model the shear mechanical impedance is expressed as a parallel combination of the mechanical impedances at very low and very high frequencies. In terms of the shear moduli these models give

Barlow-Lamb	Maxwell
$\frac{G'}{G_\infty} = \frac{4(\omega\eta/2G_\infty)^{3/2} [1 + (\omega\eta/2G_\infty)^{1/2}]}{\{[1 + (\omega\eta/2G_\infty)^{1/2}]^2 + (\omega\eta/2G_\infty)\}^2}$	$\frac{G'}{G_\infty} = \frac{\omega^2 \tau^2}{1 + \omega^2 \tau^2}$
$\frac{G''}{G_\infty} = \frac{2(\omega\eta/2G_\infty) [1 + 2(\omega\eta/2G_\infty)^{1/2}]}{\{[1 + (\omega\eta/2G_\infty)^{1/2}]^2 + (\omega\eta/2G_\infty)\}^2}$	$\frac{G''}{G_\infty} = \frac{\omega \tau}{1 + \omega^2 \tau^2}$
$\eta' = \frac{G''}{\omega}$	$\eta' = \frac{G''}{\omega}$

The variation of  $G'/G_\infty$ ,  $G''/G_\infty$  and  $\eta'/\eta$  with reduced frequency for the Barlow-Lamb model is shown in figure 5. This model fails to give correct predictions at very low frequencies.\* The variation of effective viscosity

---

\* Hutton, J. F.; and Phillips, M. C.: Analysis of the Linear Viscoelastic Relaxation of a Liquid Hydrocarbon. To be published.



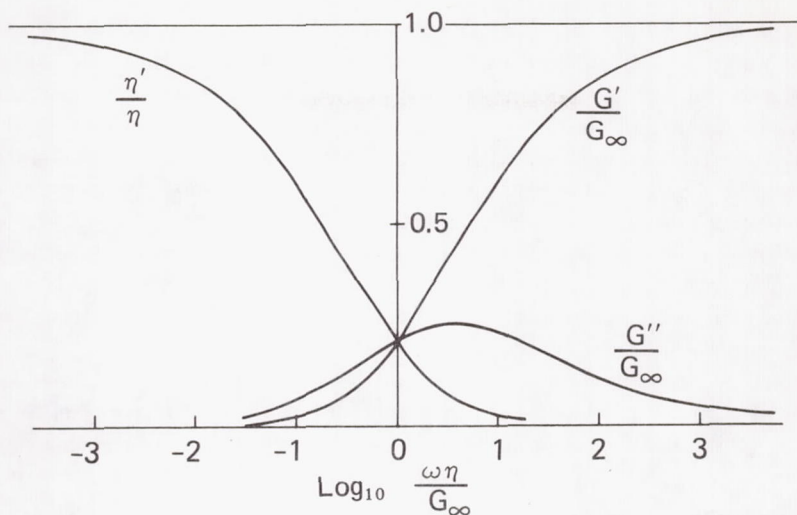


FIGURE 5.—Variation of  $G'/G_\infty$ , and  $G''/G_\infty$  and  $\eta'/\eta$  against  $\log_{10}(\omega\eta/G_\infty)$  for the liquid model of Barlow and Lamb.

with shear rate for the Ree-Eyring, Maxwell, and Barlow-Lamb models is indicated in figure 6.

An entirely different rheological model for application to the elasto-hydrodynamic situation has been proposed by Smith (ref. 8). He points out that the concept of an ideal Newtonian liquid that transmits indefinitely high shear stresses as the shear rate is increased is unrealistic

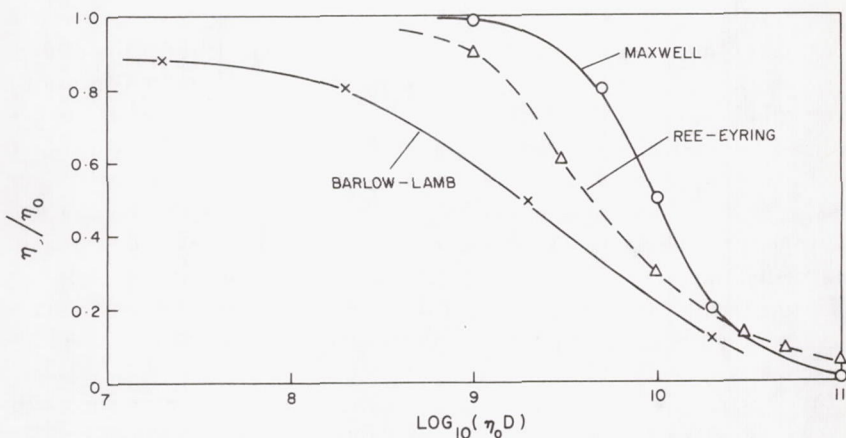


FIGURE 6.—Flow properties predicted by the Ree-Eyring, Maxwell, and Barlow-Lamb models.

when applied to real liquids. He suggests, therefore, that at high shear rates liquids exhibit a limiting shear stress analogous to the yield stress of plastic solids. It is then supposed that shear is restricted to a shear plane of molecular dimensions sandwiched between layers of essentially rigid material. The analogy with plastic solid behavior is superficial, since at yield in plastic solids shear is not, in general, restricted to one particular plane; also, rate-of-strain effects are observed. However, the thermal conditions in a thin film of liquid of low thermal conductivity are very different from those existing in the bulk deformation of a metal. In the former case temperature variations across the film tend to concentrate the shearing process at the central plane (ref. 19).

Plint (ref. 10) has adopted this hypothesis to explain his own observations of a limiting envelope of frictional traction for results obtained in a disc machine. He has suggested a molecular interpretation of the shearing process that gives approximately the required form of shear stress velocity characteristic, viz, a linear relation between  $\tau$  and (velocity)<sup>1/4</sup>.

To complete the plastic shear hypothesis, it is necessary to examine the dependence of the limiting shear stress on temperature, pressure, and shear rate. This is discussed later in this paper.

Finally there is a possibility that a phenomenon analogous to the melt fracture observed in high-molecular-weight polymers may occur. Hutton (ref. 20) has observed such an effect using polydimethylsiloxanes in a cone-and-plate viscometer and has proposed a criterion (ref. 21) that relates a fraction of the stored elastic energy to the energy required to form the fracture surface. The presence of a high hydrostatic pressure would seem to preclude the formation of a fracture surface of this kind inside an EHL contact, but it could be important in the inlet region.

#### APPLICATIONS

A detailed comparison of the way in which these various hypotheses have been applied to the available experimental data derived from disc and similar machines is complicated by two factors. First, with a somewhat restricted exception in the case of the thermal and viscoelastic approaches, no two hypotheses have been rigorously compared on the same set of experimental results. Second, in the case of the Ree-Eyring and viscoelastic models, the required data have not been available for the particular fluids used in disc machine experiments. Indeed, only in rare circumstances (ref. 6) have viscoelastic and disc machine experiments been carried out on the same liquids, and in that instance the disc machine tests did not include measurements of the frictional tractions.

We therefore consider to what extent the models have been successful in explaining the observed facts in the individual cases. We take first the pioneering work of Crook (ref. 7), who demonstrated very clearly the decrease in effective viscosity with increasing rolling speed at low sliding

speeds and concluded that for these near-isothermal conditions the observations could not be explained in terms of a Newtonian lubricant with viscosity a function of temperature and pressure. He attempted an explanation in terms of the Maxwell viscoelastic model, but found that the value of the elastic rigidity modulus necessary to explain his results was of the order of  $5 \times 10^7$  dyne/cm<sup>2</sup>, whereas the value obtained in oscillatory shear experiments for the type of mineral oil used was of the order of  $10^{10}$  dyne/cm<sup>2</sup>. He then turned his attention to the high-sliding-speed tractions and again sought an explanation in terms of a Newtonian lubricant with exponential viscosity-temperature-pressure characteristics. He was forced to conclude that this could be achieved only if the coefficients themselves were allowed to vary with the operating conditions.

More recently Plint (ref. 9) has himself arrived at the same conclusion in considering results covering a much wider range than those of Crook. He points out that for an explanation in these terms the parameter (temperature coefficient of viscosity/thermal conductivity) must vary over a range of 600/1. We still lack a full thermal treatment of the friction in an elastohydrodynamic contact, but it seems highly improbable that it will account for the observations even at high sliding speeds. Crook's analysis was a tentative one and failed to take account of the fact—pointed out by a number of workers, e.g., Tanner (ref. 15)—that the Maxwell and similar models cannot be applied in a coordinate system that maintains a fixed orientation with respect to the point of contact.

Dyson (ref. 19) has made a more detailed study of the application of viscoelastic liquid models to the elastohydrodynamic situation. For the Maxwell model, there is an equivalence between the behavior in continuous and that in oscillatory shear such that:

continuous shear

$$\frac{P'_{12}}{G_{\infty}} = \frac{Z_1}{1 + Z_1^2}$$

$$\frac{P'_{11}}{G_{\infty}} = -\frac{P'_{22}}{G_{\infty}} = \frac{Z_1^2}{1 + Z_1^2}$$

$$P'_{33} = 0$$

Where

$$Z_1 = \frac{\eta_0 D}{G_{\infty}}$$

oscillatory shear

$$G^* = G' + jG''$$

where

$$\frac{G''}{G_{\infty}} = \frac{Z_2}{1 + Z_2^2}$$

$$\frac{G'}{G_{\infty}} = \frac{Z_2^2}{1 + Z_2^2}$$

$$Z_2 = \frac{\eta_0 \omega}{G_{\infty}}$$



so that the behavior in continuous shear is equivalent to that in oscillatory shear with  $\omega$  replaced by  $D$ , and the effective viscosity in continuous shear is given by

$$\eta = \frac{G''(D)}{D} = \frac{G''(\omega)}{\omega}$$

Dyson discusses the known objections to the use of the Maxwell model which are as follows:

(1) It predicts a maximum in the shear stress as the initial viscosity,  $\eta_0$ , or the shear rate is increased (refs. 13 and 22).

(2) The predicted normal stresses are not in accord with observation (ref. 23).

(3) It is based on small strain theory whereas at stresses approaching the limiting elastic modulus,  $G_\infty$ , the strains will be large.

More important perhaps in the present context is that the variation of effective viscosity with shear rate for lubricating liquids of interest is less rapid than that predicted by the Maxwell model.

To overcome these objections Dyson proposes to use a modified form of a more general model used by Oldroyd (ref. 14). It will not be discussed in detail, but the net effect is to introduce a "shift factor" into the Maxwell equations so that equivalence is established by replacing  $\omega$  by  $KD$ . The model was applied to the results of Crook (ref. 7) and Smith (ref. 22), and a typical correlation is seen in figure 7. No data were available for the behavior in oscillatory shear of the oils used in the traction experiments, and it was necessary to use data obtained on roughly similar oils by Barlow and Lamb (ref. 17). It should be noted that although the value of  $K$  is chosen such as to reconcile the oscillatory and continuous shear results, there are restrictions on this choice that make the process less arbitrary than may appear at first sight.

The Barlow-Lamb model promises to provide a more satisfactory representation of the behavior of real liquids, but no analysis based on it has been published. However, such an analysis has been undertaken and will be published soon.\*

Perhaps the most suitable examples of the application of the Ree-Eyring model to the EHL problem are those by Bell (ref. 24) and by Bell, Kannel, and Allen (ref. 25). The first relates to the development of a Grubin-type approximation for film thickness in which the rheological equation of state (eq. (3)) is used. The analysis is complicated and a critical comparison with experiment rendered difficult by the limited access to the Ree-Eyring parameters for the liquids tested.

The film thickness predicted for a Ree-Eyring liquid is shown in

---

\* Private communication with A. Dyson.

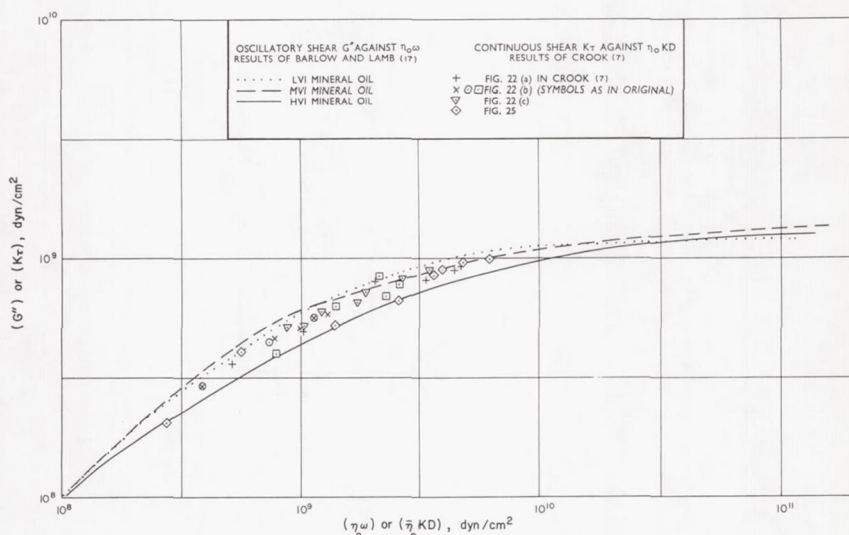


FIGURE 7.—Comparison of results of Barlow and Lamb (ref. 17) in oscillatory shear with those of Crook (ref. 7) in continuous shear (ref. 19).

figure 8 (ref. 24), which also shows the Grubin approximation for a Newtonian lubricant together with a few experimental results. It is seen that the form of Ree-Eyring representation used typically overestimates the departure from Newtonian behavior. It should also be noted that the experimental results were obtained at high rolling speeds where apparent divergence from Newtonian theory would be expected as a consequence of viscous heating in the inlet region, as mentioned earlier.

In reference 25 frictional tractions measured in a disc machine are reported which show the same characteristic behavior as those of Crook and of Smith and which lead the authors to the same general conclusion that non-Newtonian behavior of the lubricant must be invoked to obtain an explanation. Unfortunately, owing to lack of data, the authors were able to provide only a qualitative explanation in terms of the Ree-Eyring model.

The Ree-Eyring model for non-Newtonian behavior of lubricants has the great merit that it has its foundations in the molecular motions of real liquids. It has been shown to describe with considerable accuracy the behavior of some liquids in certain rheological situations. When applied to the elastohydrodynamic problem, however, it suffers from several disadvantages:

- (1) Even in the restricted form of a single hyperbolic sine function it is mathematically intractable.
- (2) In this form it predicts a decrease of effective viscosity with in-

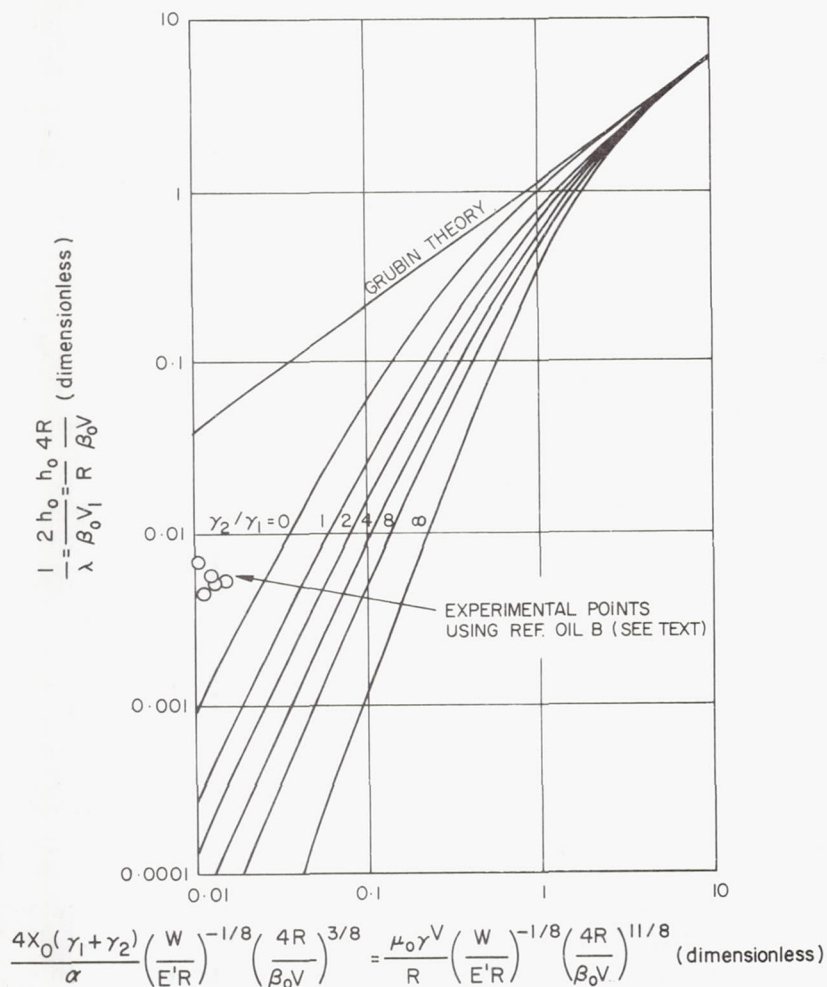


FIGURE 8.—Graph for finding film thickness of a Ree-Eyring lubricant between rolling cylinders (ref. 24).

creasing shear rate that is much more rapid than that observed in typical lubricating liquids.

(3) The material constants are complicated functions of temperature and pressure and cannot be determined without prior knowledge of the flow properties in continuous shear.

Jefferis and Johnson (ref. 26) and Johnson and Cameron (ref. 27) have produced an impressive collection of results on frictional tractions for a wide range of load and entrainment velocities. Their results confirm the



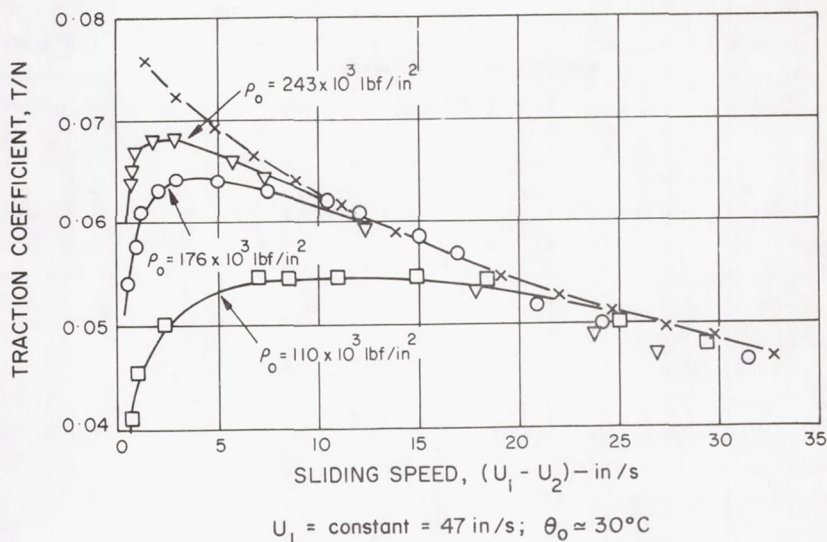


FIGURE 9.—Variation of traction with sliding speed at various contact pressures (ref. 27).

observations of earlier investigators, including the existence of an upper limit of traction coefficient at high loads. A typical set of results is seen in figure 9. Johnson and Cameron (ref. 27) have made a critical study of the plastic shear hypothesis using their own results. Smith's hypothesis is that the critical stress is a function of pressure and temperature only. Johnson and Cameron have considered this together with a second possibility, viz, that the critical stress is a function of pressure and shear rate only. For many of the results, good correlation was found with Smith's hypothesis, the critical stress,  $\tau_c$ , being related to mean Hertz pressure,  $\bar{p}$ , and the calculated temperature at the shear plane,  $\bar{\theta}_c$ , by the empirical expression

$$\frac{\tau_c}{\bar{p}} = 0.083 \left\{ \frac{\bar{p}}{139,000} \right\}^{0.2} \left\{ \frac{\theta_c}{30} \right\}^{-0.4}$$

but there was serious disagreement for results obtained at a high disc temperature.

The alternative hypothesis was not examined in detail by Johnson and Cameron. It was dismissed on the grounds that it was inconsistent with the results obtained at high rolling speeds and high disc temperatures for which the maximum traction coefficients fall appreciably below the upper bound.

They concluded that the plastic shear hypothesis is more consistent

with observation than an explanation based on a Newtonian system with frictional heating, but did not attempt a comparison with viscoelastic theory. The plastic shear hypothesis is still in an embryonic stage and so far has been studied only in direct relation to elastohydrodynamic contacts.

#### FURTHER WORK

Clearly there is a pressing need for further determinations of the flow properties of lubricants as a function of temperature, pressure, and shear rate, preferably in rheological system of simple geometry to reduce the difficulties of interpretation. This statement applies to all the models considered.

Further traction measurements would be profitable. The precise way in which these should be conducted is now fairly well understood. However, they should be carried out on lubricants for which data on as many physical properties as possible are available. This applies particularly where it is required to test the Ree-Eyring and the viscoelastic models.

#### DISCUSSION

**A. Dyson (Shell Research Limited, Thornton Research Center, Chester, England)**

This contribution is essentially a summary of a paper, "Frictional Traction and Lubricant Rheology in Elastohydrodynamic Lubrication," that the discussor has submitted to The Royal Society, London, England, and that has been provisionally accepted for publication in the *Philosophical Transactions of The Society*. It is presented now with the permission of The Royal Society.

Dr. Naylor mentions the Barlow-Lamb model as promising to provide a satisfactory representation of the frictional traction exerted under certain conditions by real lubricants in elastohydrodynamic contacts and states that an analysis based on it will be published shortly. This contribution gives an outline of the analysis.

Of the alternative models suggested by Dr. Naylor as offering a possible description of experimental results in frictional traction, the viscoelastic model has the advantage that information is available from other fields in the form of numbers that may be inserted into the appropriate equations to test the applicability of the model. Other possible schemes of explanation conspicuously lack such numerical information. The Barlow-Lamb liquid is a special model of a viscoelastic fluid with certain advantages when compared with the more familiar Maxwell model.

The treatment is intended to apply to experimental conditions similar to those used by Johnson and Cameron (ref. 27), in which the surfaces were smooth on the scale of the surface roughnesses and most of the frictional traction could be regarded as arising from hydrodynamic effects. Under these circumstances many of the experimental results may be

explained quantitatively on the basis of thermal effects in a Newtonian lubricant, provided that the pressures are low (Hertzian maximum compressive pressure less than approximately 50 000 lbf in.<sup>-2</sup>). For pressures higher than about 100 000 lbf in.<sup>-2</sup>, the predicted friction is increasingly larger than the experimental friction, particularly at relatively low rolling speeds.

In a typical disc machine experiment reported by Johnson and Cameron, the mean rolling speed  $\bar{U} = \frac{1}{2}(U_1 + U_2)$  of the discs and their bulk temperatures were maintained at constant values, and the minimum film thickness was therefore constant. A typical variation of frictional traction with sliding speed is illustrated in figure 10 and may be divided into three regions as shown. Since the film thickness is constant, the traction vs sliding speed curve is equivalent to a shear stress vs shear rate curve, and the linear region seems at first sight to be Newtonian in character. However, the viscosity calculated from the slope of this linear portion decreases with increasing rolling speed, as Dr. Naylor mentions, and this is evidence of non-Newtonian effects.

Thermal effects are obviously important at high sliding speeds. The increase in the temperature of the lubricant in the median plane of the oil film, compared with that of the bounding metal surfaces, may easily be calculated in an approximate manner. It is found that the increases are

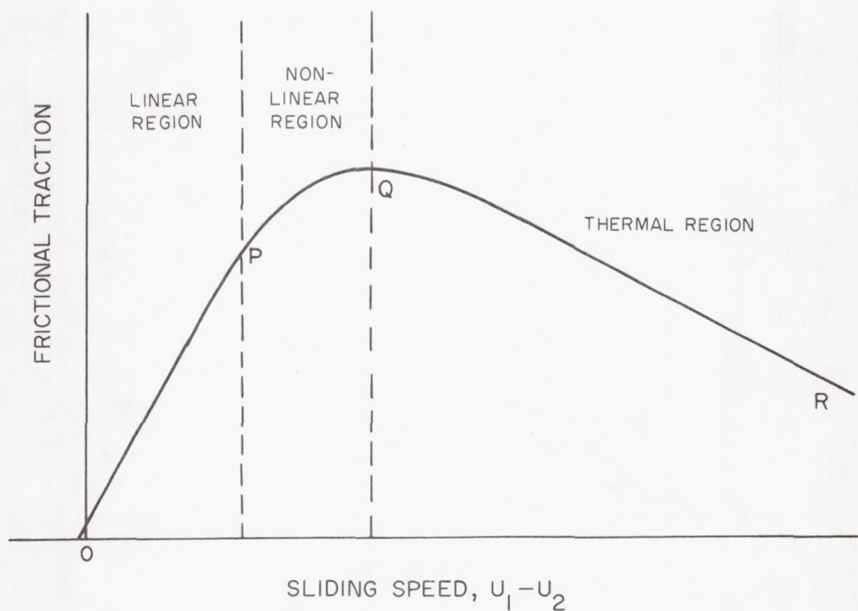


FIGURE 10.—Variation of frictional traction with sliding speed.



typically about  $5^{\circ}$  to  $10^{\circ}$  C at the maximum of curves of the type shown in figure 10. It is therefore plausible to attribute the descending part of the curve in figure 10 to thermal effects, and this region has been named the thermal region. The curvature of the nonlinear region cannot be attributed entirely to thermal effects. This is further evidence of non-Newtonian behavior.

When a number of such curves as that shown in figure 10 are obtained by varying the conditions of operation such as rolling speed, disc temperature, and load, a limiting "ceiling" curve is found beyond which no curve will penetrate. This behavior is illustrated in figure 9 of Dr. Naylor's paper and is quite inconsistent with thermal effects in a Newtonian fluid with a viscosity increasing approximately exponentially with pressure. Thus all three regions of the traction sliding speed curves furnish evidence of non-Newtonian behavior. These regions will be discussed in turn.

*The linear region of the traction vs sliding speed curve.*—The decrease of apparent viscosity with increasing rolling speed, i.e., with decreasing residence time of the lubricant in the conjunction, was first noted by Crook (ref. 7), who attempted an explanation in terms of the time required for the shear stress in a Maxwell viscoelastic liquid suddenly exposed to a constant shear rate,  $D$ , to build up from its initial value of zero. Crook had to assume values of the limiting shear modulus at high frequencies,

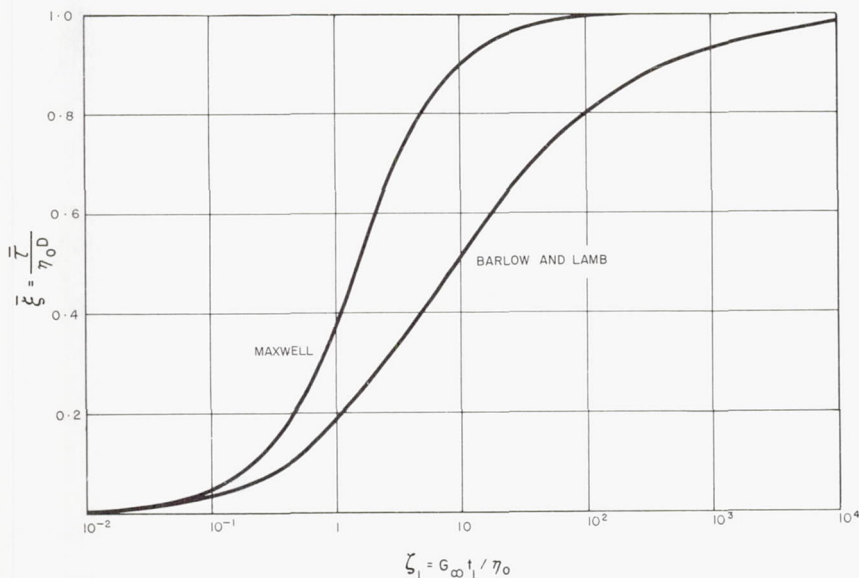


FIGURE 11.—Comparison of response of different viscoelastic model liquids in continuous shear at low shear rates.

$G_\infty$ , as low as  $5MN\ m^{-2}$  ( $5 \times 10^7\ \text{dyn cm}^{-2}$ ) whereas the experimental values found in oscillatory shear were several tenths of  $1\ GN\ m^{-2}$  ( $10^{10}\ \text{dyn cm}^{-2}$ ).

The behavior of a Barlow-Lamb (refs. 18 and 28) fluid under these conditions must be discussed now. This fluid has been defined with reference to oscillatory shear, as Dr. Naylor has already mentioned, and the behavior of a viscoelastic fluid in continuous shear at low shear rates (small compared with the reciprocal of the characteristic time of the fluid) is rigorously related to its behavior in oscillatory shear by the linear theory of viscoelasticity. The behavior of the Barlow-Lamb model is compared with that of the Maxwell model in figure 11. The unstressed liquid with a steady-state viscosity of  $\eta_0$  is suddenly exposed to a shear rate  $D$  for a time  $t_1$ , and the mean shear stress  $\bar{\tau}$  observed during that time is calculated. The main difference between the two models lies in the range of values of the dimensionless time  $\zeta_1 = G_\infty t_1/\eta_0$ , between the behavior at long times, when the dimensionless shear stress  $\bar{\xi} = \bar{\tau}/\eta_0 D$  approaches unity, and the behavior at short times, when it approaches  $\zeta_1/2$ . Thus a change in  $\bar{\xi}$  from 0.8 to 0.2 requires about 1 decade of  $\zeta_1$  for the Maxwell liquid and 2 for the Barlow-Lamb liquid.

To assess the implications of the results of figure 11, it is convenient to work in terms of the apparent viscosity

$$\eta_a = \bar{\tau}/D$$

The variation in the apparent viscosity with rolling speed may now be worked out. If the true viscosity,  $\eta_0$ , is constant, the variation of  $\eta_a$  with  $\bar{U}$  depends on the parameter

$$\zeta_1 = \frac{2bG_\infty}{\eta_0\bar{U}}$$

where  $2b$  is the total width of the Hertzian conjunction zone. Most of the variation in  $\zeta_1$  will be caused by variations in  $\eta_0$ , and the variations in the other parameters will be relatively small. Therefore, fixed values may be assumed for the other parameters, viz,

$$b = 10^{-4}m\ (10^{-2}\ \text{cm},\ 3.94 \times 10^{-3}\ \text{in.})$$

$$G_\infty = 1GN\ m^{-2}\ (10^{10}\ \text{dyn cm}^{-2};\ 145\ 000\ \text{lbf in.}^{-2})$$

With the above assumptions the variation of the apparent viscosity,  $\eta_a$ , with rolling speed,  $\bar{U}$ , is shown in figure 12 for a Barlow-Lamb liquid. There will be some experimental lower limit to the rolling speed,  $\bar{U}$ , and from figure 15 of reference 27, it seems that this limit is about  $0.2\ m\ s^{-1}$  ( $8\ \text{in. s}^{-1}$ ). This limit is shown in figure 12.

Johnson and Cameron extrapolate their apparent viscosities back to zero rolling speed, and from figure 12 it is seen that this would be easy

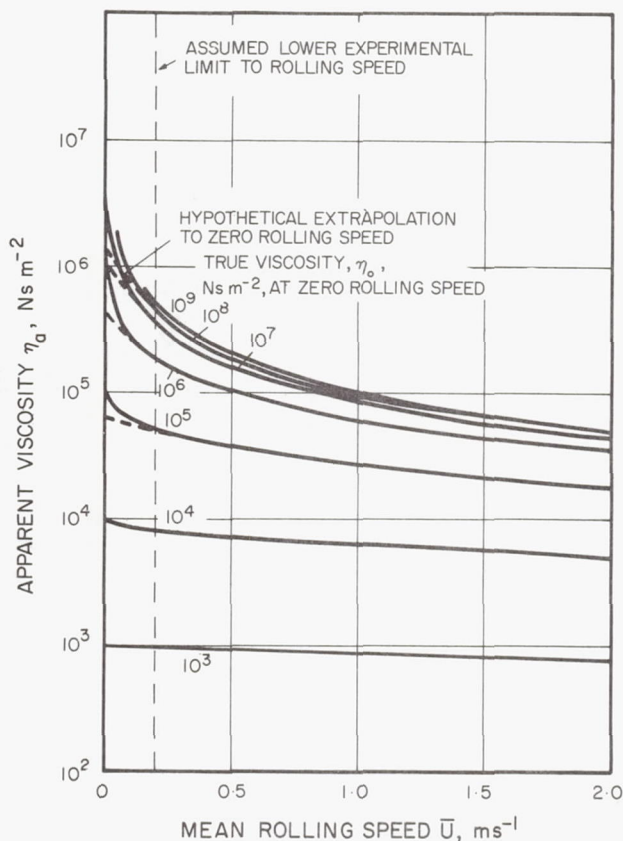


FIGURE 12.—Variation of apparent viscosity with rolling speed for Barlow-Lamb viscoelastic liquid.

for a Barlow-Lamb liquid with a low true viscosity,  $\eta_o$ , of, say,  $10^3 \text{ N s m}^{-2}$  ( $10^4 P$ ), but that it would be impossible for liquids with high true viscosities,  $\eta_o$ , say  $10^9 \text{ N s m}^{-2}$  ( $10^{10} P$ ). The dotted lines show plausible attempts at extrapolation to zero rolling speed of hypothetical sets of experimental determination of the frictional traction, limited to a lowest rolling speed of  $0.2 \text{ m s}^{-1}$ . The results of such extrapolation are shown in figure 13, in which the extrapolated apparent viscosity,  $\eta_a$ , is plotted against the true viscosity,  $\eta_o$ , for both a Barlow-Lamb and a Maxwell liquid. There is a break in the curve at a viscosity of approximately  $3 \times 10^4 \text{ N s m}^{-2}$  ( $3 \times 10^5 P$ ) for the Barlow-Lamb liquid and at  $3 \times 10^5 \text{ N s m}^{-2}$  ( $3 \times 10^6 P$ ) for the Maxwell liquid. The horizontal scale of figure 13 is linear in the true viscosity,  $\eta_o$ , and therefore approximately linear in pressure. Johnson and Cameron plot the logarithm of their extrapolated apparent viscosities



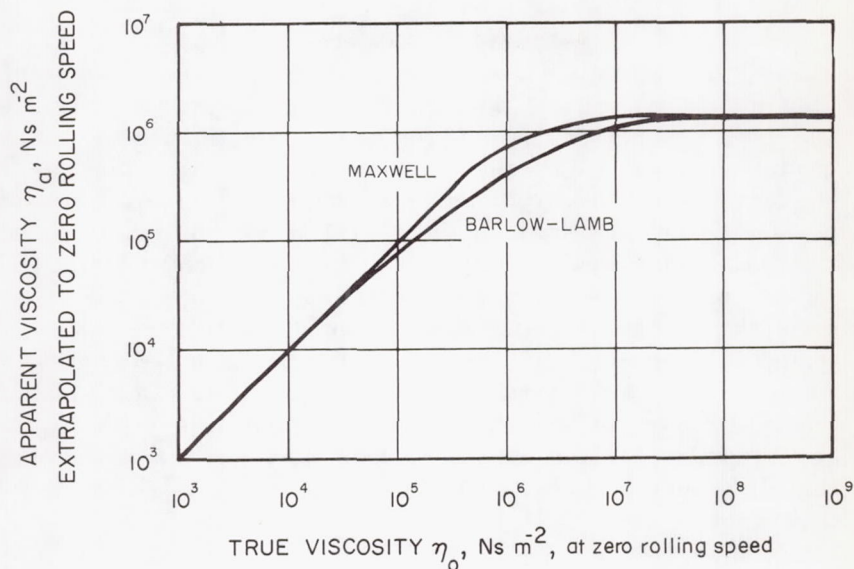


FIGURE 13.—Results of extrapolation to zero rolling speed of apparent viscosities of Barlow-Lamb and of Maxwell viscoelastic liquids.

against pressure and find a distinct break in the curves at viscosities of about  $10^5 P$ . Thus the theory for the Barlow-Lamb liquid shows considerably better agreement with the experimental results than does that for the Maxwell liquid.

An alternative way of testing the results is to examine the variation of the coefficient of friction with slide/roll ratio at low sliding speeds. Johnson and Cameron show that the results plotted in this way fall on a single line at different rolling speeds but at constant pressure and temperature. Now the mean shear stress is

$$\bar{\tau} = \eta_o D \bar{\xi}(\xi_1) \quad (4)$$

where

$$D = \frac{U_1 - U_2}{h_o} \quad (5)$$

$(U_1 - U_2)$  being the sliding speed and  $h_o$  the film thickness in the conjunction. Now there are experimental and theoretical grounds for the assumption that

$$h_o = C_1(\bar{U})^{0.7} \quad (6)$$

and equations (4), (5), and (6) then result in

$$\tau = C_2 \frac{U_1 - U_2}{\bar{U}} \frac{\bar{\xi}(\zeta_1)}{\zeta_1^{0.3}}$$

where  $C_1$  and  $C_2$  are constants for constant temperature and pressure. Thus the observation that the plots of friction against slide/roll ratio,  $(U_1 - U_2)/\bar{U}$ , coincide for different rolling speeds requires that the quantity  $\bar{\xi}(\zeta_1)/\zeta_1^{0.7}$  be approximately constant for different rolling speeds, i.e., for different values of  $\zeta_1$ . Figure 14 shows that this is a better approximation for a Barlow-Lamb liquid than it is for a Maxwell liquid.

*The nonlinear (ascending) region of the traction vs sliding speed curve.*—In the nonlinear (ascending) portion of the traction vs sliding speed curve, the shear rate,  $D$ , is not small compared with the reciprocal of the characteristic time,  $\eta_0/G_\infty$ , of the liquid, and the linear theory of viscoelasticity no longer applies. There is then no rigorous mathematical relation between the behavior of a viscoelastic liquid in oscillatory and that in continuous shear, but there is some experimental and theoretical support for an analogy between the two fields. This analogy may be put in the form

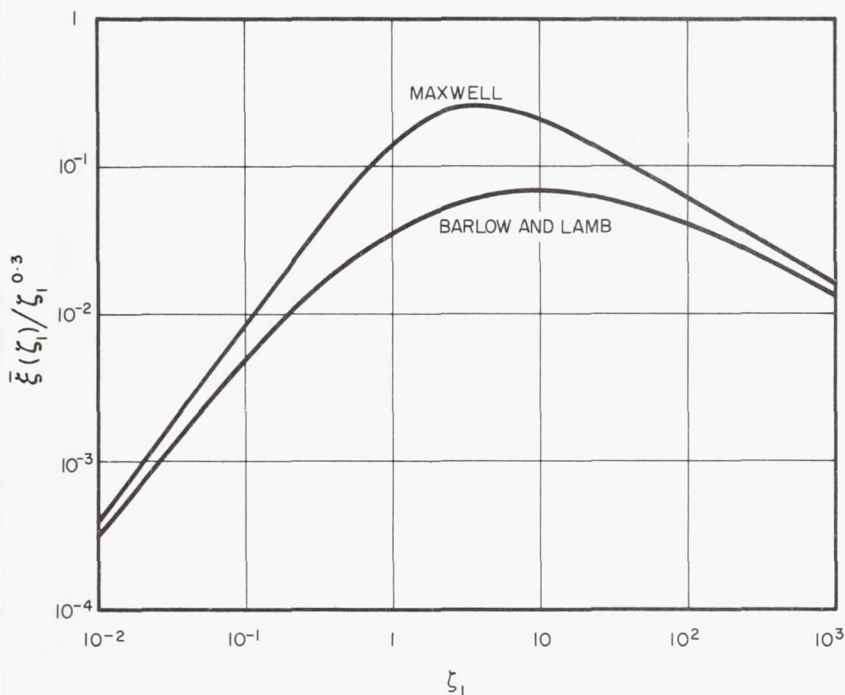


FIGURE 14.—Variation of  $\bar{\xi}(\zeta_1)/\zeta_1^{0.3}$  with  $\zeta_1$  for a Barlow-Lamb liquid and for a Maxwell liquid.

$$\tau(D) \doteq \frac{1}{K} G''(\omega) \quad (7)$$

$$\omega = KD \quad (8)$$

where  $\tau$  is the shear stress in continuous shear at a shear rate  $D$ ;  $G''$  is the imaginary part of the complex shear modulus in oscillatory shear at an angular frequency  $\omega$ ; and  $K$  is an arbitrary constant.

This analogy was used by the discussor in a previous paper (ref. 19) in which some results obtained by F. W. Smith (ref. 22) were analyzed. The curves of frictional traction against sliding speed were fitted to the curves of  $G''$  against  $\omega$  obtained for a number of different mineral oils by Barlow and Lamb (ref. 17). It was assumed that  $G_\infty$  was constant, and it was found necessary to allow  $K$  to increase with increasing temperature and to decrease with increasing pressure in an apparently arbitrary manner. It is now known, however, that  $G_\infty$  for mineral oils decreases with increasing temperature (ref. 29) and increases with increasing pressure (refs. 30 and 31). Now since

$$\frac{G''(\omega)}{G_\infty} = \phi(\eta_0 \omega / G_\infty)$$

then equations (7) and (8) give

$$\tau = \frac{G_\infty}{K} \phi(\eta_0 KD / G_\infty) \quad (9)$$

where the function  $\phi$  has been given in Naylor's paper.

Thus  $G_\infty$  and  $K$  enter into the expression for the shear stress only as their ratio  $G_\infty/K$ , and the apparent variation of  $K$ , with  $G_\infty$  assumed constant, could have been caused by a variation of  $G_\infty$ , with  $K$  constant. To test this possibility we need information about the variation of  $G_\infty$  with pressure and temperature. At atmospheric pressure, Hutton (ref. 29) found for HVI mineral oil

$$\frac{1}{G_\infty} = 2.52 + 0.024 \theta$$

where  $G_\infty$  is expressed in units of  $GN\ m^{-2}$  ( $10^{10}$  dyn  $cm^{-2}$ ; 145 000 lbf in. $^{-2}$ ) and  $\theta$  is the temperature in  $^{\circ}C$ . The variation of  $G_\infty$  of three liquids with pressure at constant temperature has been found (refs. 30 and 31) to be linear with a slope of approximately 3. It therefore seems reasonable to expect a correlation between  $G_\infty/K$  and an expression of the form

$$\frac{3\bar{p}}{2.52 + 0.024 \theta}$$

where  $\bar{p}$  is the mean Hertzian pressure. In figure 15 such a relation is



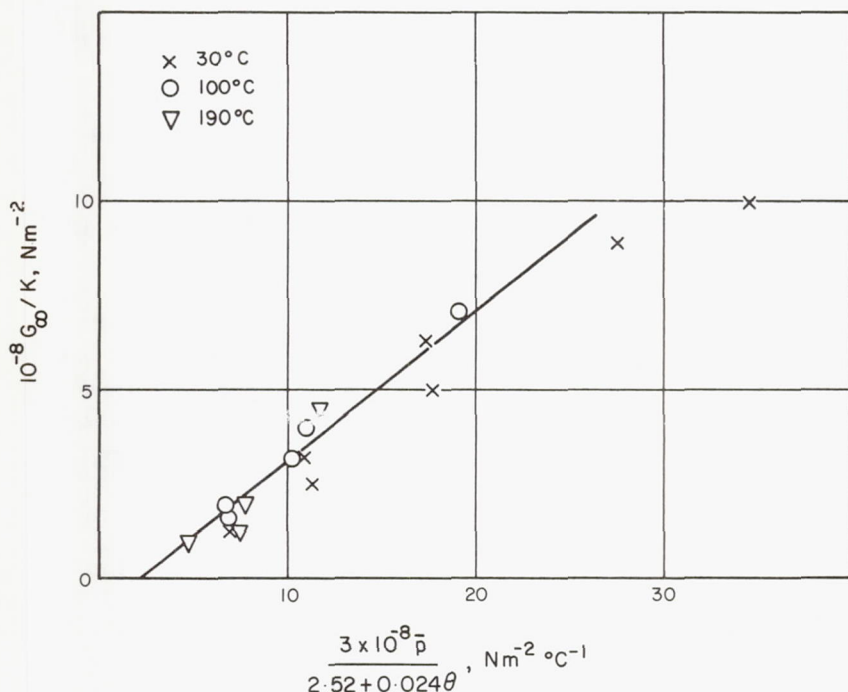


FIGURE 15.—Correlation for  $G_\infty/K$  from the results of Smith (ref. 22).

shown, deduced from the discussor's analysis of Smith's results. Although there is some scatter in figure 15, the assumed form of the variation between shear modulus, temperature, and pressure seems to be consistent with the results of the analysis. A constant value of  $K$  of 7.5 seems to fit the observations.

The viscoelastic analogy may also be tested against Plint's observation (ref. 10) that the frictional traction in this nonlinear (ascending) region varies linearly with the logarithm of the sliding speed. This is understandable on the basis of the viscoelastic analogy since both the Barlow-Lamb and the Maxwell fluid show extended regions in which  $G''$  varies linearly with  $\log \omega$ . These relations are shown in figure 16, in which equation (9) has been used to convert the variables from those used in oscillatory shear to those used in continuous shear. The agreement between the shape of these curves and Plint's results may be tested by a comparison between the ratios of the slopes of the linear portions of the curves and the maximum. Thus Plint's results (ref. 32) for

$$\frac{d(f/f_{\max})}{d \ln(U_1 - U_2)}$$

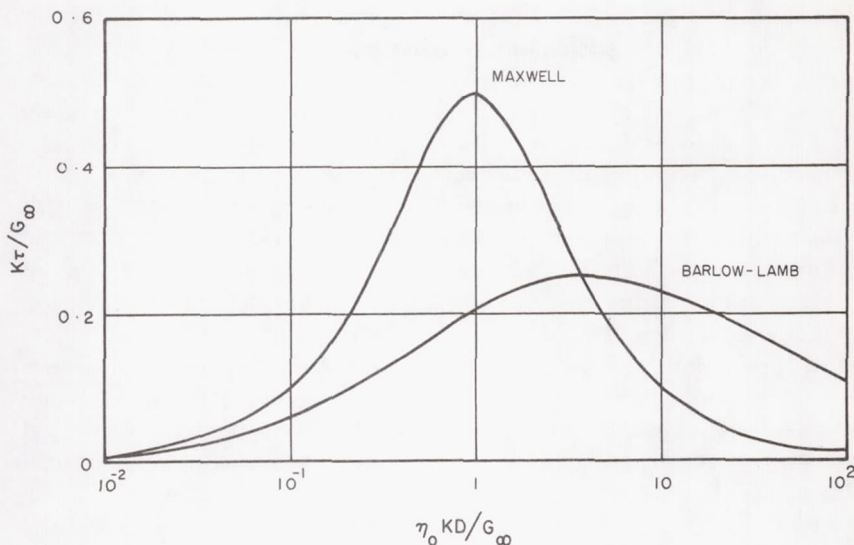


FIGURE 16.—Prediction of shear stress for viscoelastic liquids.

where  $f$  is the coefficient of friction, range from 0.27 to 0.38 with one exception, and the mean value is 0.35.

From figure 16, the value of

$$\frac{d(\tau/\tau_{\max})}{d \ln(\eta_0 D/G_\infty)}$$

is 0.28 for the Barlow-Lamb liquid and 0.50 for the Maxwell liquid. Once again the agreement is better for the Barlow-Lamb liquid than for the Maxwell liquid. The most remarkable conclusion is that this very crude version of the viscoelastic analysis gives the slope of the linear portions of the curves of friction against logarithm of sliding speed, correct to within 25 percent.

*The thermal (descending) region of the traction vs sliding speed curve.*—Johnson and Cameron (ref. 27) show that, under their experimental conditions and in the region under discussion, the shear stress in the lubricant, calculated from the observed coefficient of friction and the estimated theoretical film thickness, depends on the temperature of the median plane of the lubricant film in the conjunction. They used an HVI mineral oil, similar to Smith's although of rather higher viscosity. It is now assumed that the relation between  $G_\infty/K$ , pressure, and temperature, shown in figure 15 and deduced from an analysis of Smith's results, applies also to the oil used by Johnson and Cameron. Then the latter

authors' relation between shear stress and temperature in the median plane may be used to test the viscoelastic hypothesis, with no further disposable constants. Such a test is shown in figure 17, and the agreement is unexpectedly good. This agreement suggests that the two regions of the friction vs sliding speed curve on either side of the maximum represent one continuous relation, controlled by the same viscoelastic properties of the lubricant. This concept is opposed to that of a failure of the liquid in shear, proposed by Plint (ref. 10) and by Smith (ref. 22).

*Conclusion.*—The observed relation between frictional traction and sliding speed has been divided into three regions. Each of these regions has been examined in turn, and a reasonable quantitative agreement has been found in each region between the experimental results and theoretical predictions based on the viscoelastic properties of the lubricant, as deduced from observations made in other fields. The Barlow-Lamb model of a viscoelastic liquid shows better agreement between theoretical predictions and experimental results than does the more conventional Maxwell model. Although many gaps in our knowledge remain to be filled, the results of the analysis suggest that the viscoelastic properties of the

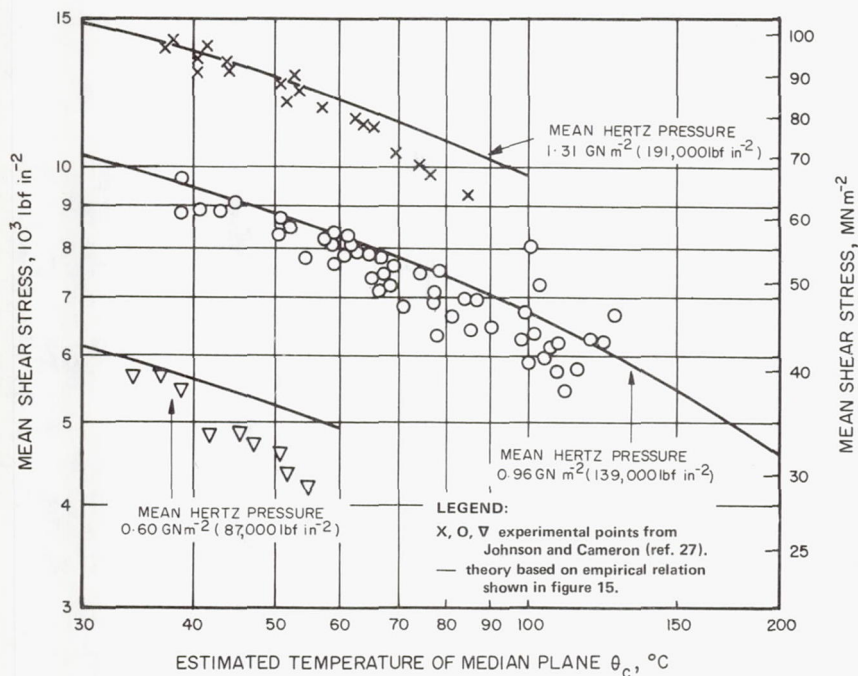


FIGURE 17.—Variation of mean shear stress with temperature and pressure.



lubricant are of major importance in determining the coefficient of friction under the conditions used in the experiments described.

#### LECTURER'S CLOSURE

I am grateful to Mr. Dyson for his brief exposition on the application of the Barlow-Lamb model of a viscoelastic liquid to the elastohydrodynamic system. Although more data are required before final conclusions can be reached, this approach does appear to provide a satisfactory explanation of the observed results in terms of known continuum behaviour of liquids. It seems to me, therefore, to offer greater prospects of advance than does the postulation of fracture mechanisms, the existence of which has not been demonstrated.

#### NOMENCLATURE

$D$	rate of shear
$G^*$	complex shear modulus of liquid
$G'$	real part of complex modulus
$G''$	imaginary part of complex modulus
$G_\infty$	shear modulus of liquid at infinite frequency
$K$	a shift factor
$P$	pressure
$S$	shear strain
$T$	temperature
$m, n$	exponents in power-law fluid equation
$\bar{p}$	mean Hertz pressure
$t$	time
$u$	velocity component
$y$	coordinate normal to flow direction
$x, X$	parameters in the Ree-Eyring fluid
$\alpha$	
$\beta$	
$\eta$	viscosity
$\eta_0$	initial viscosity
$\eta'$	apparent viscosity at frequency $\omega$
$\theta$	temperature on shear plane at critical shear stress
$\lambda$	relaxation time
$\tau$	shear stress
$\tau_c$	critical shear stress
$\omega$	angular frequency

#### REFERENCES

1. CROOK, A. W.: The Lubrication of Rollers. II. Film Thickness with Relation to Viscosity and Speed. Phil. Trans. Roy. Soc. (London), vol. A254, 1961, p. 223.

2. SIBLEY, L. B.; AND ORCUTT, F. K.: Elastohydrodynamic Lubrication of Rolling-Contact Surfaces. *ASLE Trans.*, vol. 4, 1961, p. 234.
3. DYSON, A.; NAYLOR, H.; AND WILSON, A. R.: The Measurement of Oil Film Thickness in Elastohydrodynamic Contacts. *Symp. on Elastohydrodynamic Lubrication*, *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1965-66, p. 119.
4. CHENG, H. S.; AND ORCUTT, F. K.: A Correlation between the Theoretical and Experimental Results on the Elastohydrodynamic Lubrication of Rolling and Sliding Contacts. *Symp. on Elastohydrodynamic Lubrication*, *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1965-66, p. 111.
5. ARCHARD, J. F.; AND KIRK, M. T.: Film Thickness for a Range of Lubricants under Severe Stress. *J. Mech. Engrg. Sci.*, vol. 6, no. 1, 1964, p. 101.
6. DYSON, A.; AND WILSON, A. R.: Film Thicknesses in Elastohydrodynamic Lubrication by Silicone Fluids. *Fourth Lubrication and Wear Conv.*, *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3K, 1965-66, p. 97.
7. CROOK, A. W.: The Lubrication of Rollers. IV. Measurements of Friction and Effective Viscosity. *Phil. Trans. Roy. Soc. (London)*, vol. A255, 1963, p. 281.
8. SMITH F. W.: The Effect of Temperature in Concentrated Contact Lubrication. *ASLE Trans.*, vol. 5, 1962, p. 142.
9. PLINT, M. A.: Some Recent Research on the Perbury Variable-Speed Gear. *Symp. on Elastohydrodynamic Lubrication*, *Proc. Inst. Mech. Engrs.*, vol. 180, pt. 3B, 1965-66, p. 225.
10. PLINT, M. A.: Traction in Elastohydrodynamic Contacts. *Proc. Inst. Mech. Engrs.*, vol. 182, 1967-68, p. 300.
11. MARTIN, B.: Some Analytical Solutions for Viscometric Flows of Power Law Fluids. *Int. J. Nonlinear Mech.*, vol. 2, no. 4, 1967, p. 285.
12. REE, T.; AND EYRING, H.: Theory on Non-Newtonian Flow. II. Solution System of High Polymers. *J. Appl. Phys.*, vol. 26, 1955, p. 800.
13. FROMM, H.: Laminar Flow of Newtonian and Maxwellian Liquids. *Z. Angew. Math. Mech.*, vol. 28, 1948, p. 43.
14. OLDROYD, J. G.: Non-Newtonian Effects in Steady Motion of Some Idealized Elastico-Viscous Liquids. *Proc. Roy. Soc. (London)*, vol. A245, 1958, p. 278.
15. TANNER, R. I.: Full Film Lubrication Theory for a Maxwell Fluid. *Int. J. Mech. Sci.*, vol. 1, 1960, p. 206.
16. MASON, W. P.: Measurement of the Viscosity and Shear Elasticity of Liquids by Means of a Torsionally Vibrating Crystal. *Trans. ASME*, vol. 69, 1947, p. 359.
17. BARLOW, A. J.; AND LAMB, J.: The Viscoelastic Behavior of Lubricating Oils under Cyclic Shearing Stress. *Proc. Roy. Soc. (London)*, vol. A253, 1959, p. 52.
18. BARLOW, A. J.; ERGINSAY, A.; AND LAMB, J.: Viscoelastic Relaxation of Supercooled Liquids. II. *Proc. Roy. Soc. (London)*, vol. A298, 1967, p. 481.
19. DYSON, A.: Flow Properties of Mineral Oils in Elastohydrodynamic Lubrication. *Phil. Trans. Roy. Soc. (London)*, vol. A258, 1965, p. 66.
20. HUTTON, J. F.: Fracture of Liquids in Shear: The Effect of Size and Shape. *Proc. Roy. Soc. (London)*, vol. A287, 1965, p. 222.
21. HUTTON, J. F.: Fracture of Liquids in Shear. *Nature*, vol. 200, 1963, p. 646.
22. SMITH, F. W.: Lubricant Behavior in Concentrated Contacts—The Effect of Temperature. *Rep. Nat. Res. Lab.*, MP-17, Ottawa, Canada, 1960.
23. JOBLING, A.; AND ROBERTS, J. E.: Goniometry of Flow and Rupture. *Rheology—Theory and Applications*, F. Eirich, ed., vol. 2, ch. 13, pp. 503-535.
24. BELL, J. C.: Lubrication of Rolling Surfaces by a Ree-Eyring Fluid. *ASLE Trans.*, vol. 5, 1962, p. 160.
25. BELL, J. C.; KANNEL, J. W.; AND ALLEN, C. M.: The Rheological Behavior of the Lubricant in the Contact Zone of a Rolling Contact System. *Trans. ASME*, J.

- Basic Engrg., vol. 86D, 1964, p. 423.
26. JEFFERIS, J. A.; AND JOHNSON, K. L.: Sliding Friction Between Lubricated Rollers. *Proc. Inst. Mech. Engrs.*, vol. 182, 1967-68, p. 281.
  27. JOHNSON, K. L.; AND CAMERON, R.: Shear Behaviour of Elastohydrodynamic Oil Films at High Rolling Contact Pressures. *Proc. Inst. Mech. Engrs.*, vol. 182, 1967-68, p. 307.
  28. BARLOW, A. J.; LAMB, J.; MATHESON, A. J.; PADMINI, P. R. K. L.; AND RICHTER, J.: Viscoelastic Relaxation of Supercooled Liquids. I. *Proc. Roy. Soc. (London)*, vol. 298, 1967, p. 467.
  29. HUTTON, J. F.: Viscoelastic Relaxation Spectra of Lubricating Oils and Their Component Fractions. *Proc. Roy. Soc. (London)*, vol. A304, 1968, p. 65.
  30. PURSLEY, W. C.: A Study of the Viscoelastic Properties of Liquids in the Mega-Hertz Frequency Range and in the Pressure Range of 1 to 40,000 Bars. Ph.D. Thesis, Univ. of Glasgow, 1968.
  31. SLIE, W. M.; AND MADIGOSKY, W. M.: Pressure Dependence of the Elastic Moduli of Liquid Glycerol. *J. Chem. Phys.*, vol. 48, 1968, p. 2810.
  32. PLINT, M. A.: Traction in Elastohydrodynamic Contacts. Ph.D. Thesis, Univ. of London, 1967.



**Page intentionally left blank**

# The Mechanism of Contact Fatigue

W. E. LITTMANN

The Timken Roller Bearing Company  
Canton, Ohio

The mechanism of contact fatigue can be understood in terms of several sources of stress concentration within the macroscopic contact stress field. In the subsurface region, hard nonmetallic oxide inclusions are the predominant stress raisers at which fatigue cracks originate. When cracks reach the surface, the lubricant and atmosphere exert a modifying influence on propagation behavior. Surface-origin fatigue damage can be localized at a variety of surface flaws, or general surface damage can occur when elastohydrodynamic (EHD) film thickness is small compared to surface texture at relatively high levels of contact stress. Surface-origin cracks propagate rapidly because the lubricant and atmosphere affect them as soon as they appear. Lubricant viscosity and composition both affect the origin and propagation of contact fatigue cracks, whether from surface or subsurface origins. It is useful to classify contact fatigue damage according to a system that is related to the origin, propagation mode, and macroscopic appearance of advanced stages of damage. Material and environmental factors, especially temperature and lubrication, coupled with load, contact geometry, surface microtopography, and EHD film thickness are the primary variables that combine to control the predominant mode of contact fatigue damage and the useful life that can be obtained in a given situation.

CONTACT FATIGUE has been of particular interest because it is a unique set of phenomena resulting from the interaction of relatively hard materials with load, contact geometry, lubrication, and other environmental effects. The extreme variations of apparent contact fatigue strength we sometimes encounter are the best evidence that we are observing the effects of several variables that defy accurate measurement and sometimes combine in unexpected alliances to produce astonishing endurance or perplexing infant mortality.

The early published work on contact fatigue utilized mechanical engineering and elastic stress analysis with little emphasis on materials. The powerful influence of hardness on contact fatigue strength was discovered and exploited in the earliest ball bearing technology and employed the tool steels with which European toolmakers were familiar. In gears and

in line-contact bearings, the more complex shapes were machined from softer steels of lower carbon content and surface hardened by carburizing when cold rolling was not sufficient to provide adequate load capacity. Needed improvements in performance of contact stress devices were evolutionary in character since the precision of machining equipment was clearly related to the precision of the bearings that supported their rotating parts.

Until the relatively recent developments in steelmaking practice and the demands of high-temperature and high-speed operation in aerospace applications spurred material variations in ball and roller bearings, most published studies of material effects in contact fatigue were related to gears and cams. In the past 2 decades, experimental and theoretical studies of contact fatigue have disclosed that the lubricant composition and contact surface materials properties are important as well as microtopography of surfaces, elastohydrodynamics, etc., so that it is clear we are dealing with a complex problem.

#### STATISTICAL ASPECTS OF THE PROBLEM

A major concern of makers and users of rolling-contact bearings was the wide scatter in fatigue life observed when identically made bearings were tested under closely controlled laboratory test conditions. The probability approach to load ratings was manifested in the  $B_{10}$  life or "rating life" universally used in bearing applications. A major boost to experimental efforts was provided by the acceptance of Weibull statistics to describe the distribution of bearing or contact rig test specimen lives. Equally important were the statistical confidence tools described by Johnson (ref. 1) and others.

If a very large group of identical bearings are tested or used under a given load, speed, and environment, the lives of individual bearings cannot be predicted. The distribution of life is more predictable and the catalog-rated  $B_{10}$  life is that which 90 percent or more of a very large population of a given bearing will exceed under given conditions. It is expected that as many as 10 percent of all bearings will sustain fatigue damage before the  $B_{10}$  life and still meet the catalog rating. If greater reliability is required, a lower probability of failure must be used as basis for bearing application.

Weibull (ref. 2) has provided a mathematical expression for the skewed distribution function which best describes the lives of individual bearings in a large population tested under given conditions. In cumulative form the expression is

$$S = e^{-(L/a)^m} \quad (1)$$

transformed to linear form,



$$\log \log \frac{1}{S} = m \log L - m \log a \quad (2)$$

where  $S$  = probability of survival (estimate of percent population surviving time  $L$ )

$L$  = individual bearing life

$a$  = characteristic life (a life value at which 63.2 percent bearings have failed)

$m$  = Weibull slope (dispersion parameter)

By plotting the results of individual bearing life tests on Weibull paper or by a computer fit of life data to the Weibull distribution, one obtains estimates of the  $L_{10}$  or  $L_{50}$  lives (10 and 50 percent failed in a population) for the sample representing a given population of bearings with a specific geometry, processing, and testing history. Thus  $L_{10}$  is an estimate of a finite population life, and  $B_{10}$  is a catalog-rated life for an essentially infinite population comprising greater variations of life due to materials, processing, and application conditions.

The early observation of fatigue cracks and microstructural alterations in the region of maximum shear stresses beneath the contact surfaces of ball bearings and the similar appearance of spalls resulting from contact fatigue in ball bearings and rig tests led to a widely held belief that all rolling-contact fatigue cracks initiated at oxide inclusions beneath the surface. It seemed logical that the variation of the composition, size, frequency, and location of inclusions with respect to the well-known subsurface shear stress distribution would account for most of the observed scatter in bearing life.

Indeed, the advent of vacuum-melted steels quickly provided evidence that a significantly lower oxygen content, manifested in the steel as a lower frequency-severity distribution of oxide inclusions, did provide significant improvements in bearing life. However, the observation that vacuum-melted steel consistently gave no improvement in life for some bearings while showing the expected increase in others furnished a clue to the more complex nature of contact fatigue (ref. 3).

For some time there was a spirited controversy over whether contact fatigue cracks formed first at the surface or subsurface in the region of maximum shear stresses under rolling contact without gross sliding. Incipient cracks or deformation bands at inclusions (butterflies) and the clear association of computed subsurface stresses with dark etching or general deformation band (gray lines) alterations of microstructure in hardened bearing steels supported the subsurface nucleation theories. Combined rolling and sliding was believed to be the primary reason for surface-initiated cracks, seen most frequently below the pitch line of gear teeth, since maximum shear stresses are at the surface under such con-

ditions if the surface friction coefficient exceeds 0.1 (ref. 4). It is now generally recognized that contact fatigue can nucleate and propagate from several kinds of stress raisers. The location of the crack nucleus with respect to the contact surface, geometry, stress level, material strength distribution, and environmental factors modifying the contact stress distribution and influencing crack propagation combine to cause fatigue cracks that form and grow at different rates within the region subjected to the highest contact stresses. The propagation of one or more of the cracks ultimately results in spalling of material from the contact surfaces.

The competitive nature of contact fatigue as we now know it has been described in contact fatigue rigs, gears, and in several types of rolling-contact bearings. The major objective of this paper is to describe the mechanism of the initiation and propagation of the various modes of contact fatigue one may expect to encounter. Wherever a quantitative treatment can be applied, such methods will be included, but this paper will be predominantly descriptive and qualitative. A liberal sprinkling of speculative hypotheses will be included in the hope that constructive discussion will ensue and some increment of knowledge can be obtained.

#### STATEMENT OF THE PROBLEM

The first hint that the state of stress plays an important role in contact fatigue is apparent when one examines the mechanical properties of the materials used in parts subject to high contact stresses. In table 1 the uniaxial tensile and compressive strength, the bending fatigue strength, and some typical contact fatigue results are listed for typical bearing steels. The contact fatigue strength of the same materials determined in a simple rolling-contact test rig demonstrates that the ability of these materials to withstand millions of stress cycles is greatly increased by the state of stress in the rolling-contact test compared to the bending fatigue test or even uniaxial static compression.

The critical stresses in contact fatigue have been the subject of extensive study. For most practical operating conditions of nominally pure rolling contact, the distribution of pressures and shear stresses for the contact surfaces and subsurface regions are well described by Radzimovsky (ref. 7). Except for variations due to full EHD film or rolling plus sliding conditions, the orthogonal shear stress appears to undergo the maximum range. Moyar and Morrow (ref. 8) have made a comprehensive review of critical stresses in rolling contact. They also made correlation studies with fatigue under other states of stress which include the effect of compressive stresses on the planes that undergo maximum cyclic shear stresses. Theory and experiment have shown good agreement over a limited range of load and contact geometry. Greenert (ref. 9) showed that the range of orthogonal shear stress was consistently related to the contact fatigue strength of 52100 steel when it was tested at maximum Hertz pressures

TABLE 1.—*Comparison of Static, Bending Fatigue and Contact Fatigue Properties of Bearing Steels at Room Temperature (refs. 5 and 6)*

Alloy steel	Hardness, Rockwell C	Tensile strength		Compressive yield strength, ksi	Fatigue strength, ksi at $10^7$ cycles*
		Yield,	Ultimate, ksi		
52100	62	240	340	400	130
Halmo	62	310	370	410	140
M-1	62	310	370	420	130
M-50	62	295	370	340	125

Alloy steel	Hardness, Rockwell C	Contact fatigue life in NASA spin rig	
		Maximum Hertz compressive stress, ksi	$L_{10}$ cycles
Halmo	62	800**	$350 \times 10^6$
M-1	62	800	$100 \times 10^6$
M-50	62	800	$180 \times 10^6$

\* Rotary beam reversed bending fatigue tests, mechanically polished surfaces

\*\* Computed from tests run at 750 ksi

from 830 to 900 ksi with contact areas having the major axis parallel or perpendicular to the direction of rolling. The paper by Dr. H. Poritsky in this symposium and others (refs. 4, 7, 10, and 11) describe the important aspects of stress distribution under conditions of (1) pure rolling (no tractive forces tangential to the contact surfaces), (2) rolling with superimposed tractive force, and (3) rolling plus sliding.\*

The computation of nominal contact stresses from imposed load and contact geometries has been described in many publications and the validity of the results verified from photoelastic and other experimental stress analyses. While it is recognized that the true stresses which cause the initiation of contact fatigue cracks are considerably higher than the nominal stresses, the computation of these localized stresses is less easily performed (refs. 12 and 13). I shall limit myself to enumeration of the various ways in which locally intensified stresses are produced and how they are manifested in the formation of fatigue damage. Following is a list of those factors which modify the nominal contact stress distribution from that computed for homogeneous, isotropic, elastic, smooth surfaced materials in dry rolling contact.

\* Sliding is defined here as gross relative motion of the contact surfaces as distinguished from microscopic slip within the nominal contact area.



- (1) Subsurface stress raisers
  - a. Oxides and other hard, brittle inclusions
  - b. Sulfides, carbides, and other second phase particles
  - c. Grain boundaries, sub-boundaries, twins, and other dislocation arrays
- (2) Surface character
  - a. Texture
    1. Roughness
    2. Waviness
    3. Lay
  - b. Residual stresses
  - c. Surface energy level
  - d. Microstructure
  - e. Contaminants
- (3) Surface flaws
  - a. Inclusions and second phase particles
  - b. Nicks and dents including true and false brinelling
  - c. Grooves and scratches
  - d. Corrosion pits, rust, water etch
  - e. Fretting damage
  - f. Skidding damage
- (4) Discontinuities in contact geometry
  - a. End of "line" contact geometry
  - b. Debris particles in the contact area
- (5) Load distribution within the bearing
  - a. Elastic deflections
  - b. Misalignment of bearing parts
  - c. Internal clearance—bearing adjustment
- (6) Elastohydrodynamics
- (7) Tangential forces
  - a. Without gross sliding
  - b. Rolling plus sliding

#### COMPETITIVE MODES OF DAMAGE

Consideration of the modifiers of theoretical contact stress distribution leads one to appreciate the major reason for the complexity of behavior seen in contact fatigue. The study of contact fatigue is basically a study of stress concentration effects that govern fatigue crack initiation coupled with crack propagation behavior. Propagation effects have been the object of growing concern because the application of high purity steels has shifted the predominant fatigue damage origin to the surface where lubrication and other environmental effects can affect the crack from its very origin. The spectrum of subsurface and surface stress raisers can be visualized as a handicap race in which fatigue cracks originate by different

mechanisms at surface and subsurface inclusions or other discontinuities. The number of stress cycles at which the crack embryo begins to propagate depends on what we are willing to call a crack. Given a threshold of detection for the crack, we can distinguish a fatigue life for crack initiation and a life increment during which the crack propagates from the point of detection to a degree of propagation considered as failure. Once the threshold of crack detection has been passed, one can measure crack propagation rate. From a practical standpoint, it is clear that considerable propagation may have taken place on a microscopic scale before we could observe the fatigue crack, especially if the origin was subsurface. Thus we may divide the total observed fatigue life into three periods:  $L_N$ , crack nucleation;  $L_D$ , propagation to first detectable damage;  $L_p$ , measurable propagation. Life to failure is the sum of  $L_N + L_D + L_p$ . Clearly the extent of  $L_p$  that can be tolerated depends upon the crack propagation rate under prevalent operating conditions and the degree of damage that can be sustained without loss of essential functions.

#### CLASSIFICATION OF CONTACT FATIGUE DAMAGE

The usefulness of a contact fatigue damage classification scheme is primarily to facilitate communications on the subject, but it is also good for enhancing our perception of the mechanism of crack formation and growth. Intrinsic in the classification system is a hypothesis of the mechanism for each mode of crack formation that differs sufficiently from the others to require a separate classification of the resulting damage.

The most general classification of contact fatigue damage is pitting or spalling, commonly observed in all contact stress situations. Spalling is generally understood and used to indicate the removal by cracking of one or more particles from the surface of a given part. The shape of the particles or of the spalled region from which they came is usually small in depth compared to the other two principal dimensions. In British and European literature, pitting is used to mean the same thing as spalling in contact fatigue. In rolling plus sliding, as in gears and cams, pitting generally refers to surface-initiated modes of contact fatigue damage, but is sometimes used for both surface- and subsurface-initiated damage.

Some examples of contact fatigue damage classifications from the published literature are summarized in table 2. Denning and Rice (ref. 18) proposed three designations of contact fatigue damage at the same time Littmann and Moyer (ref. 3) suggested five competitive modes of contact fatigue, based upon the origin and appearance of the damage. Wren and Moyer (ref. 19) added a sixth mode. Tallian (ref. 16) distinguished between spalling and surface distress because of the difficulty in accurate classification and lack of general applicability of more elaborate

TABLE 2.—*Contact Fatigue Damage Classification Systems*

Source	Ref.	Terms used
O.E.C.D.	14	Burnish Case crushing Fatigue wear Flaking Initial pitting Pitting Spalling
Mitsuda	15	Spalling, pitting, shelling (see also his refs. 2 to 4)
Tallian	16	Spalling, surface distress
Martin and Eberhardt	17	Spalling nuclei: 1. Nonmetallic inclusions 2. Surface furrows 3. Debris dents 4. Pits
Denning and Rice	18	Modes of contact fatigue failure: 1. Subsurface pitting 2. Surface pitting 3. Case crushing
Littmann and Moyer	3	Competitive modes of contact fatigue: 1. Inclusion origin 2. Point surface origin 3. Geometric stress concentration 4. Peeling 5. Subcase fatigue (in casehardened parts)
Wren and Moyer	19	Same as Littmann and Moyer plus: 6. Transverse cracking (in through-hardened components)

systems. Martin and Eberhardt (ref. 17) have identified four types of spalling nuclei in ball bearings.

As more data become available, we should modify our terminology to incorporate new information and always subject our classification system to a test of its general usefulness. On the basis of this logic, I propose the following as a system of classification for contact fatigue damage, whether resulting from pure rolling or rolling plus sliding, and for any material or contact geometry. Repeated static contact and oscillatory rolling contact are special cases and may not be completely described within this system.

- (1) Subsurface-origin spalling (or pitting)
  - a. Nonmetallic inclusion
  - b. Unconfirmed origin
- (2) Subcase fatigue or "case crushing"



- (3) Surface origin
  - a. Point surface origin (PSO)
    - 1. Debris dent
    - 2. Handling nick
    - 3. Surface flaw (grinding furrow, etc.)
    - 4. Surface inclusion (intersecting or parallel and coincident with surface)
    - 5. Peeling plus hydraulic pressure propagation (HPP)
    - 6. Corrosion pit
  - b. General surface distress
- (4) Geometric stress concentration (GSC)
  - a. End of "line" contact
  - b. Edge of peeled area
- (5) Peeling (or superficial pitting, sometimes called "frosting" or "glazing")
- (6) Section fracture

The mechanism by which each of these modes of contact fatigue damage initiates and propagates will be described by means of examples taken principally from laboratory tests and applications of tapered roller bearings. Wherever examples of similar damage in rolling-contact bearings of different contact geometry or in gears, cams or rig test specimens appear in published literature, selected examples have been reproduced from the publication or from photographs supplied by the authors.

Because of frequent examples from the literature and the composite character of many illustrations in this paper, it was not always convenient to identify material, heat treatment, test conditions, etc. Wherever tapered roller bearing parts are shown, they are AISI 4620, modified, steel that was case-carburized and hardened to approximately 60  $R_C$  in the near surface regions and 30  $R_C$  in the supporting core material. Other materials are identified where possible. Ball bearing parts were generally through-hardened 52100 steel at approximately 62  $R_C$ , although some were tempered to 59  $R_C$  for dimensional stability. The gear examples are probably case-carburized and hardened. Where differences in materials are pertinent to the discussion, they are noted. Arrows in the figures show the direction of load approach.

#### SUBSURFACE ORIGIN DAMAGE

The elastic stress analyses of Radzimovsky (ref. 7), Fessler and Ollerton (ref. 20), and Johnson (ref. 1) leads one to expect yielding or crack initiation at some location where the maximum range of shear stresses is intensified by some discontinuity in the material subjected to repeated stress from rolling contact. Experimental evidence of Greenert (ref. 9) indicated the maximum orthogonal shear stress range to be

critical since it gave the best correlation with fatigue life for three different contact geometries. The orientation of the cracks at early stages of development, seen in post-test examination of their test specimens, was parallel to the contact surface. A summary of the results of that investigation is given in table 3. The location of subsurface cracks was at the depth of maximum orthogonal shear stress range,  $Z_o$ , for the 0.500-inch crown radius specimens and mating cylindrical driver. Some cracking was seen at less than  $Z_o$ , especially in the 0.383-inch radius specimen tests. None of the specimens examined showed cracking at the greater depth of maximum unidirectional shear stress,  $Z_s$ .

Barwell and Scott (ref. 21) reported subsurface cracks in the "area of calculated maximum shear stress" of 52100 type steel balls tested at 740 ksi maximum surface compressive stress. Scott, Loy, and Mills (ref. 22) also reported that subsurface cracks were found in rolling elements from service applications, but further details of contact geometry and stress level were not given. In both instances, however, it was specifically noted that no general microstructural changes were seen, and the authors concluded that crack nucleation was probably related to stress concentration at discontinuities in the microstructure, citing brittle nonmetallic inclusions as a likely example. There are numerous examples in the literature which show that large nonmetallic inclusions of the refractory or glassy oxide type can act as the stress raisers at which subsurface fatigue cracks are formed (refs. 23 to 26). Most of the evidence relating smaller oxide inclusions to contact fatigue is indirect, e.g., the improved life obtained with vacuum-melted or vacuum-deoxidized (with carbon) steels. The lower frequency-severity distribution of oxide inclusions in such steels compared to air-melted steels is generally recognized as the basis of their superior contact fatigue strength. Lyne and Kasak (ref. 27) have demonstrated that increased sulfur levels in 52100 steel were associated with improved fatigue life in bending and contact fatigue rig tests. They suggest that envelopment of the oxide inclusions in sulfide reduces their effective stress concentration effects in fatigue.

There is little documentation for crack nucleation at smaller inclusions (ref. 17), although there are many published photomicrographs of local microstructural changes at hard inclusions that conclusively demonstrate severe stress concentration effects at such discontinuities in the structure (refs. 3, 12, 17, 23, 25, and 28 to 34). While the "butterfly," "deformation band" or dark etching alterations of microstructure have been the subject of considerable study, there has been no clear link between such effects and the nucleation of contact fatigue cracks. Bush et al. (ref. 33) related the boundary of dark etching regions to a threshold orthogonal shear stress range in ball bearings of hardened 52100 steel. Martin et al. (refs. 17 and 28) have shown that the propagation path of contact fatigue cracks in 52100 steel at  $R_c$  59 was affected by deformation bands and



associated lenticular carbides, but they were careful to note that no correlation of failure initiation point was found with white etching areas or the depth of 45° maximum unidirectional shear stress at which general deformation bands were most numerous. Martin and Eberhardt (ref. 17) have shown that light etching deformation bands are formed in hard 52100 steel at depths that correlate with the subsurface shear stress ( $\tau_{45}$ ) distribution and also appear at a variety of stress raisers that have been shown to be the origins of contact fatigue cracks. For several surface-initiated modes of damage, an association of light etching deformation bands with surface fatigue cracks was illustrated. Almost as often, however, surface fatigue cracks were illustrated having no associated deformation bands, so the "chicken and egg question" remains a moot point. Figure 1 (from ref. 16) shows surface cracks in 52100 ball bearing raceways that have the light etching deformation bands as well as some which do not.

Schlicht (ref. 29) has related the occurrence of "butterflies" at inclusions to the distribution of shear stresses beneath the surface for a rolling-contact fatigue rig test of an unidentified steel that appears to be 52100, as shown in figure 2. He found that oxide inclusions were associated with butterflies seven times more frequently than sulfides.

The frequent observation of local plastic deformation at inclusions but few observations of fatigue cracks proceeding beyond such alterations has led me to conclude that reversed plastic deformation on the scale of the "butterfly" or "deformation band" altered microstructure is not an essential step in the formation of subsurface-inclusion-origin contact fatigue cracks (ref. 23). Some examples of intact inclusion-origin spalls that have been sectioned to disclose the point of probable fatigue crack nucleation are shown in figures 3 to 6. These examples are typical of normal inclusion-origin fatigue damage in tapered roller bearings that showed lives within the normal catalog-rated expectancy and include relatively short and long lives within the normal life distribution. In these and many other examples of inclusion-origin contact fatigue cracks, we only rarely see "butterfly" microstructure along the inclusion which was the crack nucleus. Almost always, other inclusions at depths identical with the fatigue origin show extensive altered microstructure, and often have microcracks along the boundary of the "butterfly" or within it as shown in figure 7. The crack in figure 6b was entirely subsurface and nowhere showed more "butterfly" than seen here. It is the only example of a completely subsurface crack we have found in a bearing that extends beyond the inclusion-butterfly aggregate.

It is relatively easy to study inclusion-origin spalls in case-carburized bearings because a low propagation rate often permits the discovery of intact spalls. The load distribution in line contact contributes to the low propagation rate as well as the high fracture toughness associated with



TABLE 3.—*Comparison of Subsurface Crack Depth of Maximum Unidirectional Shear Stress and Depth of Maximum Orthogonal Shear Stress Range for Tests of Greenert (ref. 9)*  
(52100 Steel,  $R_c$  63)

Roller type		Toroid crown radius, in.	Cycles to failure×10 <sup>6</sup>	Number of subsurface cracks of indicated length, in.			Crack depth		$\frac{Z_s}{a}$	$\frac{Z_o}{a}$
				< .010	.010 to .050	> .050	$d$ , in.	$\frac{d}{a}$		
Set A										
Drive			1.1	0	0	0	Surface	—	.47	.35
Z-139 Toroid*	.383	"	0	0	0	0	—	—	.47	.35
V-55 Drive		1.4	5	0	0	0	.008	.25	.47	.35
Z-124 Drive		7.1	12	0	0	0	.010	.31	.47	.35
Z-145 Drive		9.0	7	0	0	0	.006	.18	.47	.35
W-210 Toroid	.383	9.2	1	0	0	0	.006	.18	.47	.35
W-210 Drive*		"	5	0	0	0	.010	.31	.47	.35
Set B										
W-14 Drive		0.14	9	0	0	0	.015	.37	.51	.37
W-14 Toroid*	.500	"	4	0	0	0	.015	.37	.51	.37
W-176 Toroid	.500	0.58	37	2	0	0	.015	.37	.51	.37
W-176 Drive*			10	0	0	0	.015	.37	.51	.37
V-47 Drive		2.6	79	0	0	0	.010/.020	.25/.49	.51	.37
W-215 Toroid	.500	17.9	18	0	0	0	.015	.37	.51	.37
W-215 Drive*		"	19	0	2	2	.015	.37	.51	.37

## Set C

W-200 Toroid	.500	0.14	2	5	7	.015	.35	.51	.37
W-200 Drive*		"	0	0	0	—	—	.51	.37
W-191 Toroid	.500	0.88	28	0	0	.010/.020	.23/.47	.51	.37
W-191 Drive*		"	19	0	0	.010/.020	.23/.47	.51	.37
W-254 Drive		1.1	20	7	4	.010/.020	.23/.47	.51	.37

\* Unfailed roller of the pair

$Z_s$  = Depth of maximum shear stress  $\tau_{45^\circ}$

$Z_o$  = Depth of maximum orthogonal shear stress range

$a$  = Semi-axis of contact ellipse in the rolling direction

Set A—1910-lb loads max. orthogonal shear stress range = 368.5 ksi

Set B—2433-lb load; max. orthogonal shear stress range = 379.5 ksi

Set C—2857-lb load; max. orthogonal shear stress range = 400.5 ksi

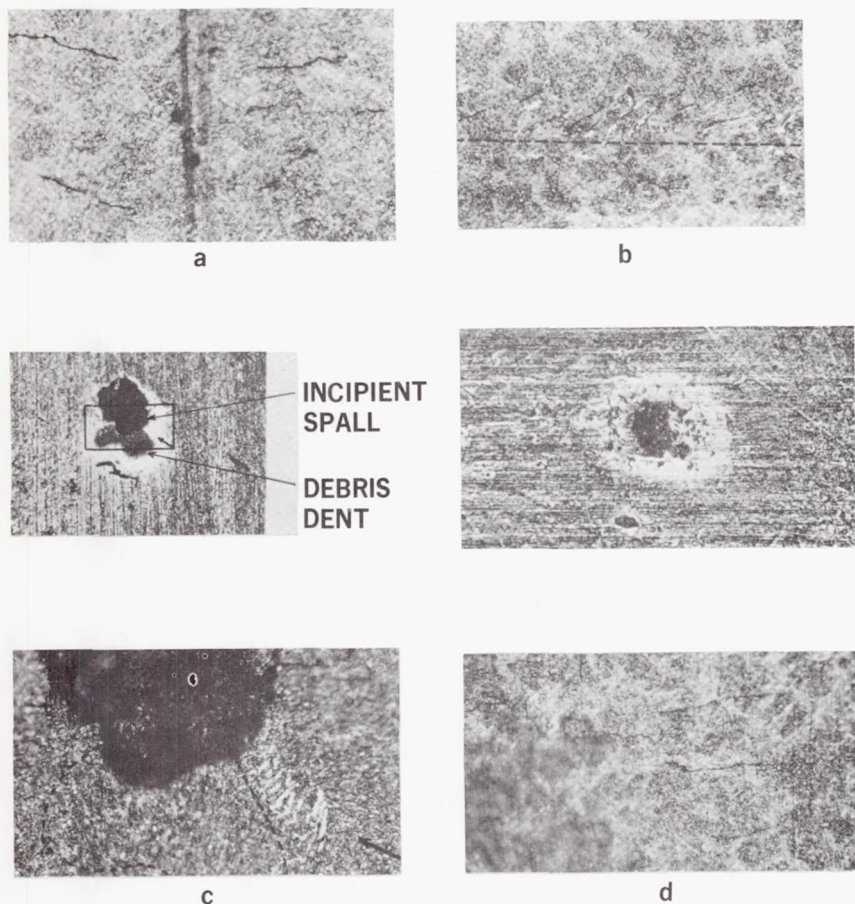


FIGURE 1.—Early stages of surface-origin contact-fatigue damage according to Martin et al. (ref. 17): (a) microcracks near a grinding furrow (fig. 26) after light metallographic polishing and etching; (b) same after polished 0.0002 in. below the original surface, note light etching deformation bands; (c) deformation bands near a debris dent; (d) cracks without deformation bands near a pit.

the levels of retained austenite (15 to 35 percent) normally present in such materials. As the fatigue cracks propagate, removal of spalled material tends to relieve contact stresses near the crack origin, and a skilled investigator can often locate the inclusion that nucleated the spall even when the extent of propagation has rendered the bearing unfit for further service. This fortunate set of circumstances has permitted us to develop a schematic diagram of crack initiation and propagation for subsurface-inclusion-origin contact fatigue damage.



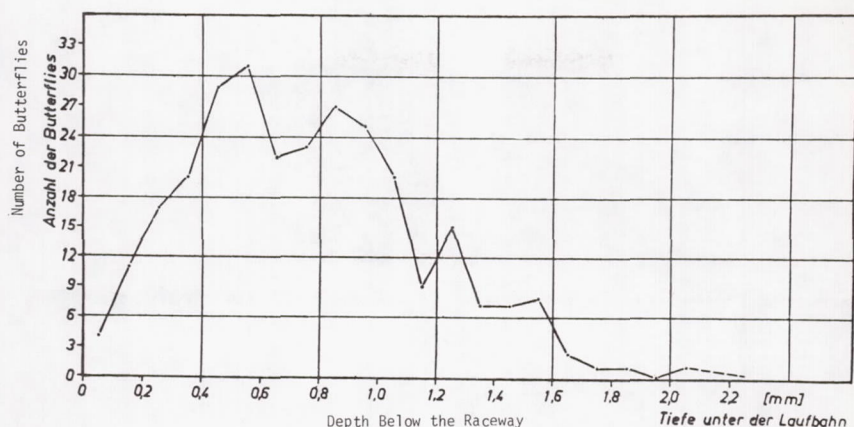


FIGURE 2.—Distribution of “butterflies” at nonmetallic inclusions as a function of depth below the contact surface of a rig test specimen according to Schlicht (ref. 29). Summation of data from specimens (probably 52100 type steel) run to  $8.74 \times 10^6$  stress cycles at 199 to 484 ksi (140 to 340 kg/mm<sup>2</sup>) maximum surface compressive stress.

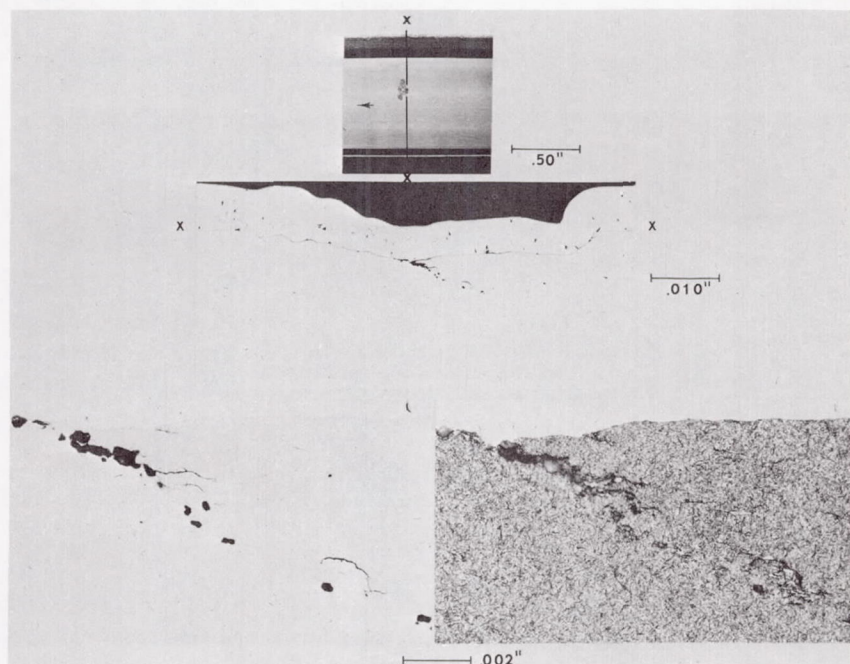


FIGURE 3.—Subsurface inclusion origin mode of damage on a tapered roller bearing cone after  $146.7 \times 10^6$  revolutions in SAE 20 oil at 166° F, maximum surface compressive stress was 315 ksi,  $h_G/\sigma_1 = 1.025$ . Depth of  $\tau_{45 \text{ max}}$  was 0.0048 in., depth of  $2\tau_{xy}$  was 0.0031 in. Nitai etch, radial plane of polish, perpendicular to direction of rolling.

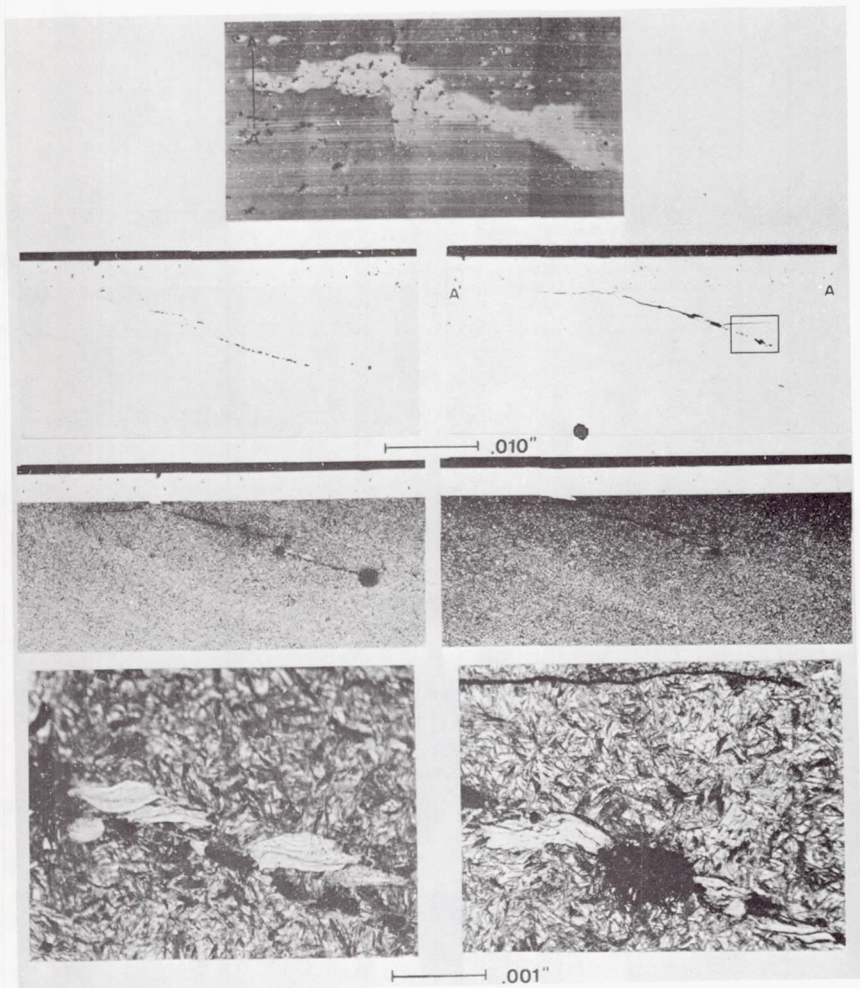


FIGURE 4.—Subsurface inclusion origin damage in a cone returned from service in an automatic transmission fluid at temperatures approaching 275° F. Note the similar tendency for the crack to propagate parallel to the surface in figure 5. Nital etch, radial plane of polish.

#### PROPAGATION OF DAMAGE FROM SUBSURFACE ORIGINS

Figure 8 shows the alternate modes of propagation observed in tapered roller bearings for spalls confirmed to have subsurface-inclusion origins. Examples of these variations in propagation behavior are illustrated in figure 9.

When circumstances favor the normal mode of propagation, cracks in the region of maximum cyclic shear stress grow faster in a direction per-

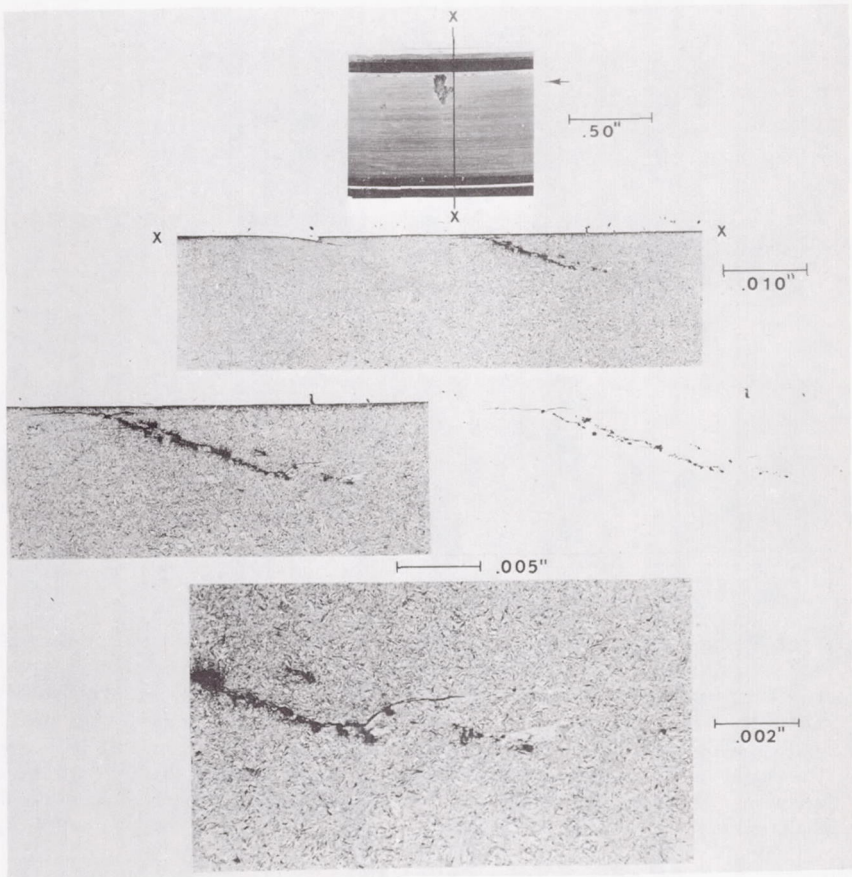


FIGURE 5.—Subsurface inclusion origin damage from laboratory tests run in Mil-L-7808F fluid at 150° F.  $h_G/\sigma_1 = 0.44$ , depth of  $\tau_{45 \text{ max.}} = 0.0054$  in., depth of  $\tau_{xy \text{ max.}} = 0.0035$  in. Here and in figure 4 the tendency to form "butterfly" alteration is seen at inclusions beyond the main crack but to a more limited extent along the cracking or not at all. Nital etch, radial plane of polish.

pendicular to the direction of rolling than in the rolling direction to form an elliptical surface that undermines the contact surface. Branching cracks running toward the contact surface eventually detach portions of the undermined surface, but so long as most of the original surface remains intact, propagation is relatively slow and harmless to bearing function. The most likely reason for lateral propagation in line-contact types of bearings is that shear stresses are not transmitted as effectively as compressive stresses are to depths beyond the propagating crack. Shear stresses are thus intensified at the lateral edges of the crack,



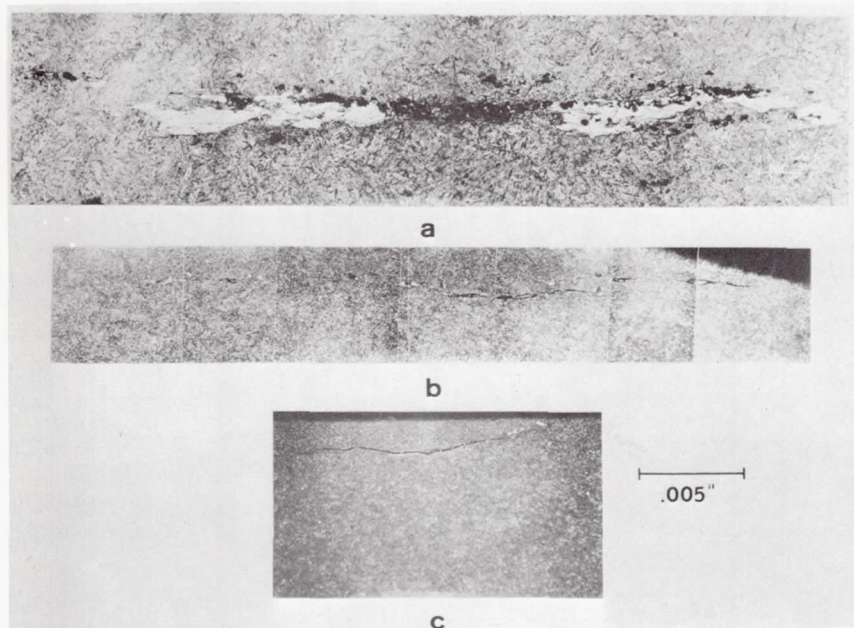


FIGURE 6.—Subsurface inclusion origin damage which did not intersect the contact surface in laboratory tests of a large bearing, 7.94-in. mean cone O.D. Cracking was located by ultrasonic inspection after test. (a) Oxide inclusions and "butterfly" alterations at 0.076 in. below the surface, after  $24.12 \times 10^6$  rev. at 260 ksi maximum surface compressive stress; (b) subsurface cracking from another cone tested under the same conditions for  $215.4 \times 10^6$  rev.; (c) same crack as (b) after 0.004-in. additional polishing. This cracking did not extend to the surface anywhere. Nital etch, radial plane of polish.

especially when the contact width in the rolling direction is less than the cracked length in that direction.

Factors that favor normal propagation are high lubricant viscosity at the bearing operating temperature and absence of chemically aggressive factors in the lubricant and atmosphere surrounding the bearing. Absence of tangential forces or of combined rolling and sliding are also believed to promote normal propagation.

Nonmetallic inclusion morphology also plays a role. Since most inclusions are not single particles but elongated clusters or stringers of particles, some of them are long enough to extend from the region of maximum cyclic shear stress to the surface. This can provide a path for crack propagation to the surface and access of the atmosphere and lubricant to the crack at the subsurface location where cyclic stresses are highest. The branching off of the propagation in the maximum cyclic shear stress region is seen in figure 4 despite the continuation of the inclusion stringer to

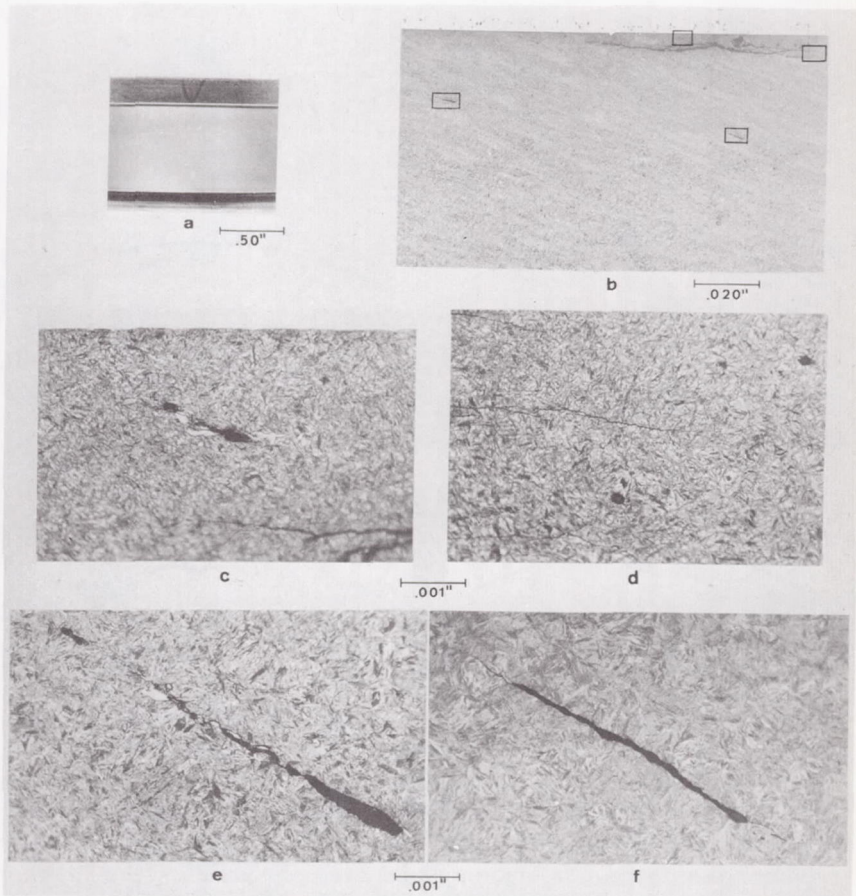


FIGURE 7.—(a) An intact subsurface inclusion origin spall on a tapered roller bearing cone. (b) Radial cross-section (perpendicular to rolling direction) of the spall in (a). Regions indicated are shown at higher magnification in (c), (d), (e), and (f). (c) Inclusion near the surface but not connected with main cracks at any point. Butterfly alteration and microcracks are evident. (d) Tip of laterally propagating crack. (e), (f) Microcracking and traces of butterfly alteration at inclusions far beneath the contact surface.

considerably greater depth. Thus the length of inclusion stringers is important as well as the size of the individual particles, and the inter-particle spacing. Since these aspects of inclusion clusters are highly variable, the chance location of heavy inclusion clustering near the contact surface is one of the primary causes of the scatter in the lives of individual bearings.

As soon as fatigue cracks from a subsurface origin reach the surface, the atmosphere and the lubricant exert a modifying influence on the crack

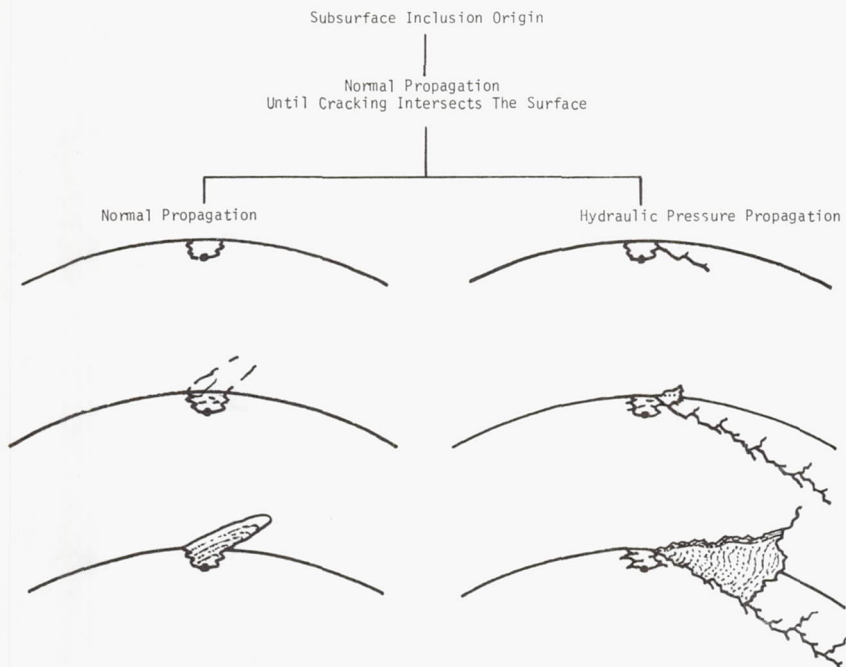


FIGURE 8.—Alternative modes of propagation for inclusion origin contact fatigue damage shown in schematic cross-sections.

propagation process (ref. 35). Evidence for this is seen in figure 9c where a normally propagating subsurface crack had clearly formed, followed by a point-surface-origin or hydraulic-pressure-propagation (HPP) type of spalling which proceeded from the trailing edge of the cracking associated with the original spall. The contrast between the smooth surface of the original slowly propagating crack and the rougher surface of the faster mode of propagation is clearly evident. Often the smooth elliptical surface of the subsurface propagating crack will show brown iron oxide from fretting of the fracture surfaces and "beach marks" at intervals when load levels changed or intermittent operation affected the crack pattern. Cross sections of such normal subsurface propagation are compared with cracks propagating rapidly under the modifying influence of the atmosphere and lubricant in figure 10. In 10A and B the radial plane of sectioning is parallel to the rolling direction, but in 10C the plane of sectioning is radial (perpendicular to the rolling direction) to show the inclusion origin of a spall that otherwise looks like PSO. As soon as the lubricant and/or atmosphere have access to the rapidly propagating edges of the crack, mostly in the rolling direction from the origin, the crack morphology is modified. Branching cracks appear, and the appearance of



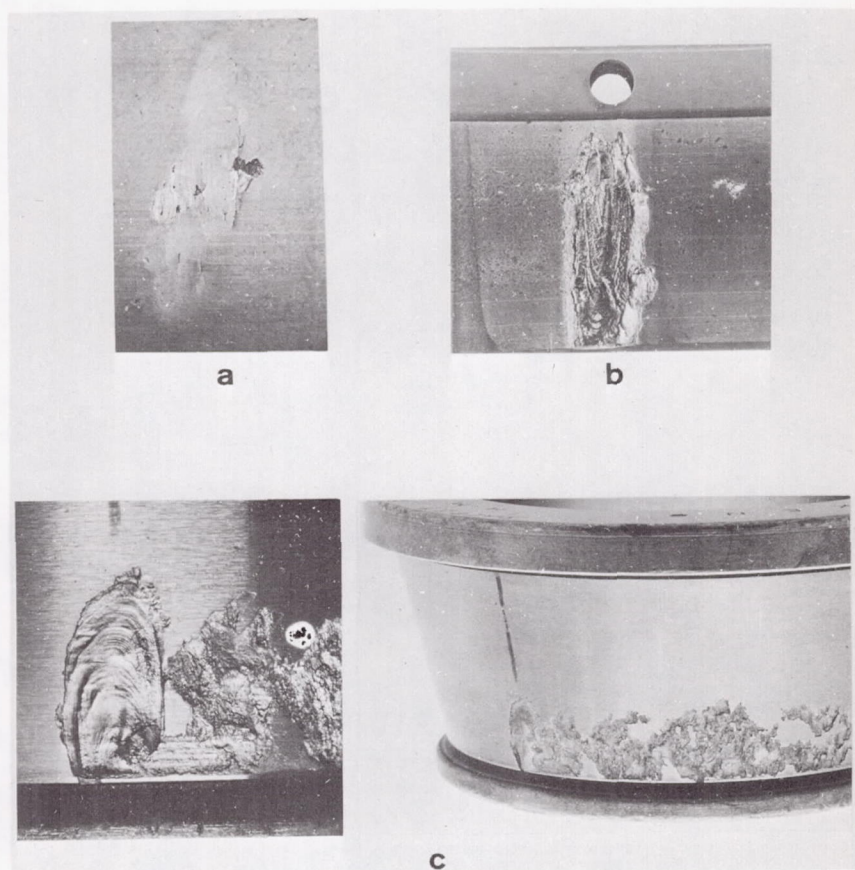


FIGURE 9.—Examples of propagation modes for inclusion origin contact fatigue damage. (a) Intact subsurface inclusion origin spall with normal lateral propagation (ref. 23), (b) advanced stages of damage after normal subsurface lateral propagation on a large cup from a steel rolling mill application, (c) normal lateral propagation from a subsurface inclusion origin near the small diameter end of contact on a heavily loaded cone which ran in Mil-L-23699 at 180° F in a helicopter application. See also figure 10.

the fracture surface is more brittle. In contrast the slower initial subsurface propagation shows a more singular crack with little tendency to branching and is more suggestive of shear than tensile forces acting at the crack tip.

Fiber direction effects on contact fatigue life are well known (refs. 30, 36, and 37). These effects are most likely the result of rapid interaction of the crack propagation process with the lubricant and atmosphere when inclusion stringers intersect the surface.

In tapered roller bearings, many inner and outer races are made from

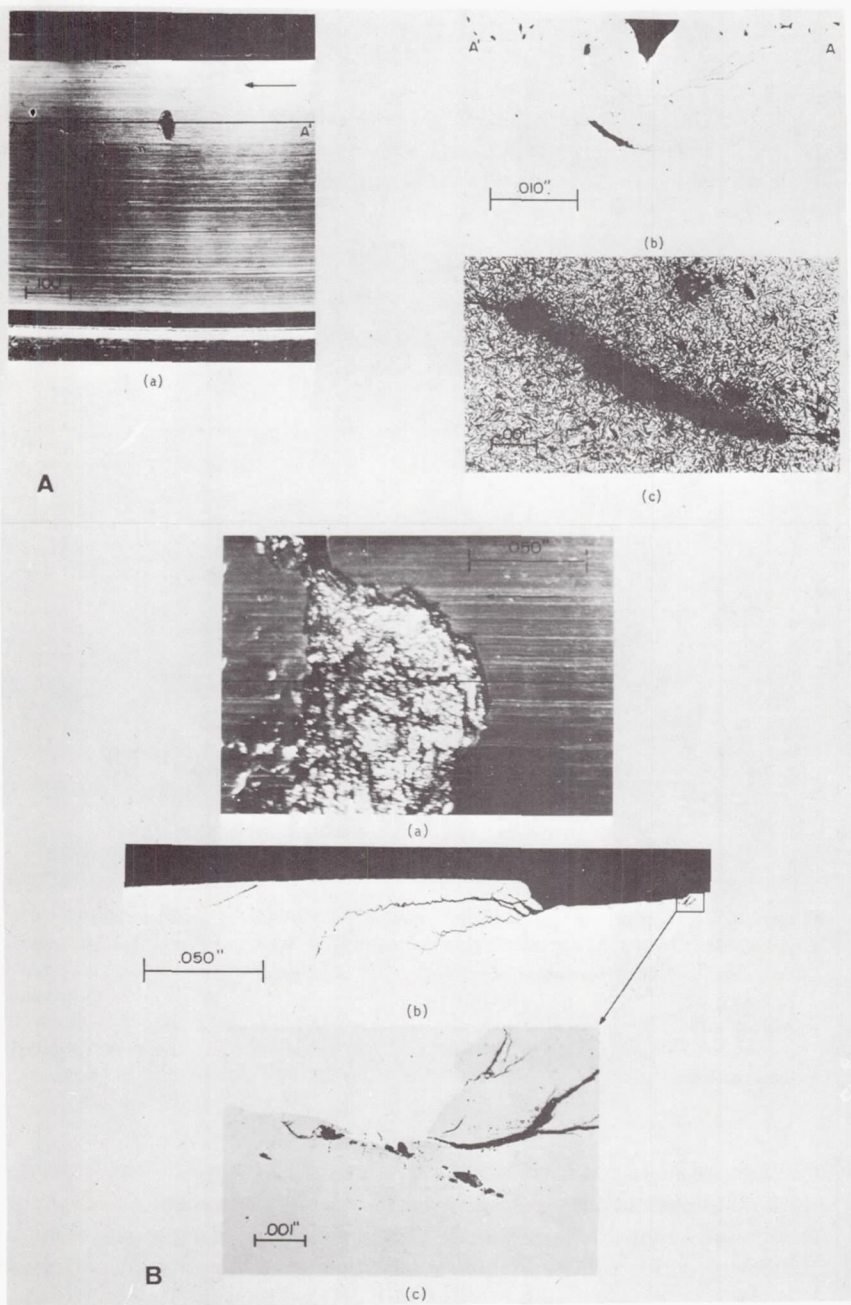


FIGURE 10.—(A) Inclusion origin with normal lateral propagation (ref. 35), operation in SAE 20 oil at 150° F, transverse plane of section, above. (B) Inclusion origin with hydraulic pressure propagation (ref. 35).

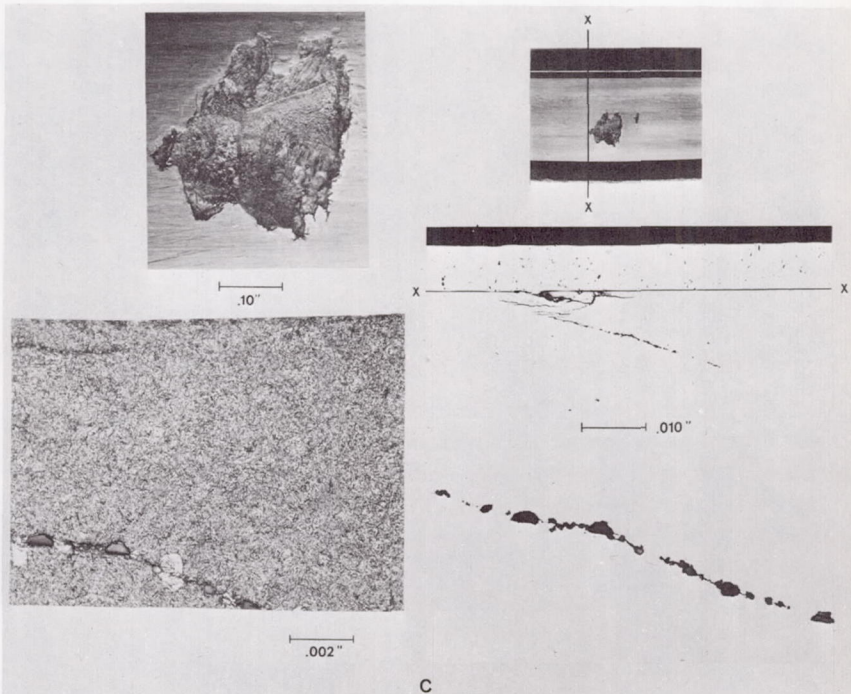


FIGURE 10 (Concluded).—(C) Inclusion origin with HPP, operation in SAE 20 oil at 166° F, radial plane of section.

seamless tubing while the rollers are headed from wire or bar. The resulting fiber directions intersect the raceway at a compound angle in the cone or cup but are essentially parallel to the roller surface. Thus an inclusion exposed on the roller surface acts more as a surface stress raiser than in a cone or cup. This leads to a somewhat inconsistent classification of an inclusion-origin spall on a roller if we normally think of an inclusion origin as located subsurface. As seen in figure 11, inclusion-origin spalls on rollers can propagate much like any other point-surface-initiated mode of damage. In any bearing or other contact surface with fiber direction controlled to be parallel to the raceway, similar effects should be seen when surface-origin damage is prevalent.

There may be instances where a chemically aggressive lubricant of low viscosity will cause rapid crack propagation from an inclusion that intersects the surface, especially if some sliding or tangential force is present, tending to promote surface rather than subsurface initiation of fatigue damage. If postmortem examination under such circumstances reveals an inclusion at the spall origin, the spall should be classified as surface inclusion origin.



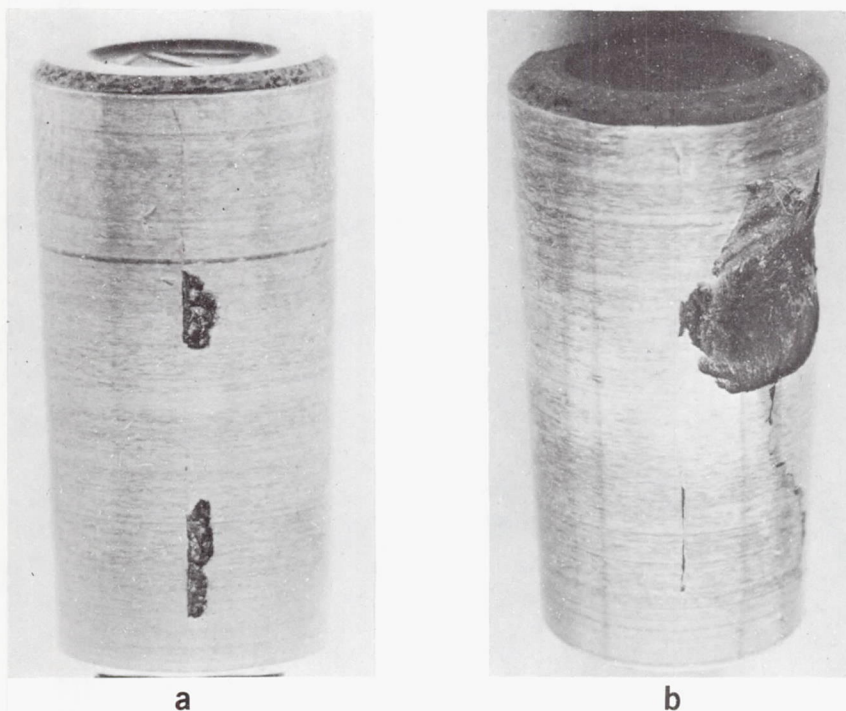


FIGURE 11.—Surface inclusion origin damage: (a) with normal lateral propagation, (b) with HPP from the same test as figure 10C. Operation in SAE 20 oil at 166° F.

#### EFFECT OF LUBRICATION ON PROPAGATION FROM INCLUSION ORIGINS

The importance of the lubricant as a modifier of inclusion-origin fatigue propagation is illustrated by the following data. Bearing life tests were run recently at The Timken Roller Bearing Company Laboratories to evaluate the effect of variations in hot deformation processing in the manufacture of bearing cones and cups. A secondary study of fatigue propagation revealed some interesting effects. All parts were made from a single heat of steel and equal numbers of the hot work variations were included in the tests so that the differences due to such variations did not affect the results of the propagation studies described here. To accelerate the testing, half of the tests were conducted in Mil-L-7808F synthetic lubricant and half in SAE 20 petroleum oil. Details of the test conditions are given in table 4.

In all tests the bearings were photographed to show the first damage ( $L_D$ ) detected by an electromagnetic vibration detection device which automatically stopped the test. Additional running was manually monitored, and the damaged part or parts were rephotographed after each

TABLE 4.—*Test Conditions for Comparison of Propagation in Mil-L-7808F and SAE 20 Oil\**

Condition	Mil-L-7808F	SAE 20
Speed (rpm)	1800	2700
Temperature (outlet oil) °F	120–135	140–155
EHD film thickness (Grubin) ( $\mu$ in.)	4	17
Composite surface roughness		
Cone—roller ( $\mu$ in.)	11.5	11.5
Cup—roller ( $\mu$ in.)	8	8
Contact stress—maximum surface compressive stress (ksi)	301	301
Bearing size		
Mean cone diameter (in.)	3.31	3.31
Mean roller diameter (in.)	0.40	0.40
Minimum propagation life increment (cone revs.)	$0.4 \times 10^6$	$1.3 \times 10^6$

\* The test bearings were tapered roller bearings and were made from 4619 (modified) steel, case carburized, hardened (R<sub>C</sub> 60) and finished in the same manner as standard product except for modifications of hot deformation processing. Tests run under radial load in 5-in. O.D. Timken bearing life test machines.

life increment ( $L_p$ ). The number of such increments between  $L_D$  and the completion of  $L_p$  ranged from 1 to 16, with most of them in the range of 4 to 14. Many of the tests were run one or more increments past the 0.01-in.<sup>2</sup> spall size used as the laboratory criterion of failure, mainly to determine whether differences in the mode of propagation could be associated with the lubricants used during initiation and propagation. In an effort to discover whether the propagation behavior was uniquely related to the lubricant, several equivalent "first-damage" spalls formed in each lubricant were propagated in the other lubricant. The results are summarized in table 5 and figure 12.

The life increment from first detected damage to 0.01 in.<sup>2</sup> for each test group is compared in table 5 as Weibull estimates of  $L_{10}$  and  $L_{50}$  with the related slope (scatter parameter) and correlation coefficient (goodness of fit) for each set of results. It was found that the propagation rate was 25 times greater in the synthetic fluid than in the petroleum oil, but in the two-fluid tests the propagation rate was intermediate between the single fluid results. Propagation in SAE 20 was more rapid when initial damage was in the synthetic. Propagation in the synthetic fluid was slower when the initial damage was in the petroleum oil.

Figure 12 shows that the appearance of the damage was remarkably similar in both fluids and in the interchanged lubricant tests, indicating a similar mode of propagation despite the large differences in propagation rate from the single-fluid tests. Figure 13 illustrates the details of propagation behavior for several of these tests. The competitive and statistical

TABLE 5.—*Effect of Lubricant on Propagation of Inclusion-Origin Contact Fatigue Damage*

Test lubricant		Weibull estimates of $L_p$				
To First Damage, $L_N + L_D$	Propagation To 0.01-in. <sup>2</sup> Spall, $L_p$	Propagation From First Damage Detected To 0.01-in. <sup>2</sup> Spall				
		No. Tested	$L_{10}$ , 10 <sup>6</sup> Rev.	$L_{50}$ , 10 <sup>6</sup> Rev.	Slope, $m^*$	Correlation Coefficient, $R^{**}$
SAE 20	SAE 20	6	19.3	50.8	1.94	.89
Mil-L-7808F	SAE 20	6	2.9	16.2	1.09	.93
SAE 20	Mil-L-7808F	7	1.4	14.5	0.80	.98
Mil-L-7808F	Mil-L-7808F	14	0.2	2.0	0.78	.91

\* The slope  $m$  is a measure of life variation within the sample, a low value indicates greater life variation.

\*\* The correlation coefficient is a measure of how well the data are represented by the Weibull distribution, i.e., deviation and scatter from a straight line on a Weibull plot. 1.00 indicates perfect linearity and no scatter of data from the line.

nature of contact fatigue damage is evident from one of the examples since a new spall origin appeared and propagated more rapidly than the original one being studied. Such behavior was more frequently observed when propagation was in the synthetic fluid. Nearly all of the spalls studied were of the subsurface-inclusion-origin type with some point-surface-origin modes appearing occasionally in the Mil-L-7808F tests. Standard vacuum-degassed steel was used, and surface texture was controlled to minimize surface-origin damage under the EHD film conditions used.

The large difference in propagation rate of the single-fluid tests is consistent with previously published work (ref. 35) and is most likely related to viscosity and chemical differences between the two lubricants used. The  $L_{50}$  life estimates for each hot work variation tested in Mil-L-7808F ranged from 6.5 to 33 percent of  $L_{50}$  for duplicate tests in SAE 20 oil. The results of the interchanged lubricant tests described above are probably best explained by a chemical effect in which the effective components of the two lubricants affecting propagation behavior remained effective despite extreme dilution in the other fluid. No attempt was made to remove the last traces of the previous fluid from the bearings in the exchanged-fluid tests. At each inspection the bearings were rinsed in kerosene and the excess blown off with air. After inspection and before replacing in the life test machines, the bearings were dipped into the



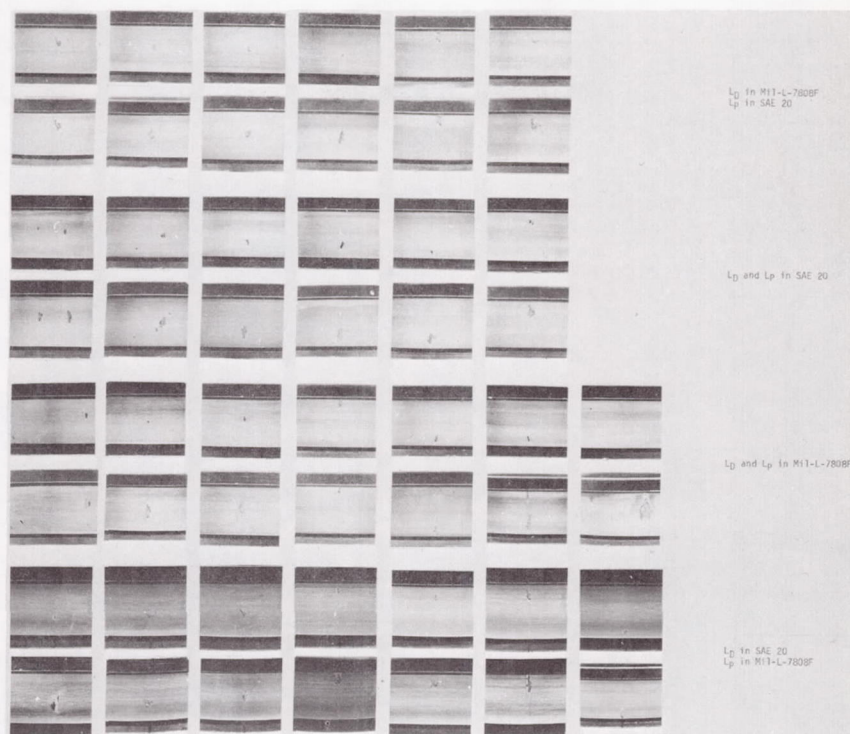


FIGURE 12.—Appearance of spalls at  $L_D$  first damage, detected by vibration, and at 0.01-in.<sup>2</sup> spalled area, used as laboratory criterion of failure. Tests under conditions given in table 4.

appropriate test oil. Thus some residue of the previous lubricant could have been retained in the cracks, but was greatly diluted in the next life increment, and successively diluted further in each subsequent inspection and test cycle. It is far more likely that a carryover chemical effect occurred than a persistence of any viscosity difference in the interchanged-fluid tests.

Additional studies on these bearings will be made by metallographic and fractographic methods in an attempt to gain a better understanding of the observed effects. For this discussion, it is sufficient to say that the life of tapered roller bearings can be greatly affected by lubricant composition, even when the dominant damage mode is subsurface inclusion origin, and propagation is relatively normal. One might be justified in suspecting that some of the conflicting results obtained by different investigators in studies of lubrication effects on bearing life could be traced to undetected differences in mode of fatigue damage origin and propagation.

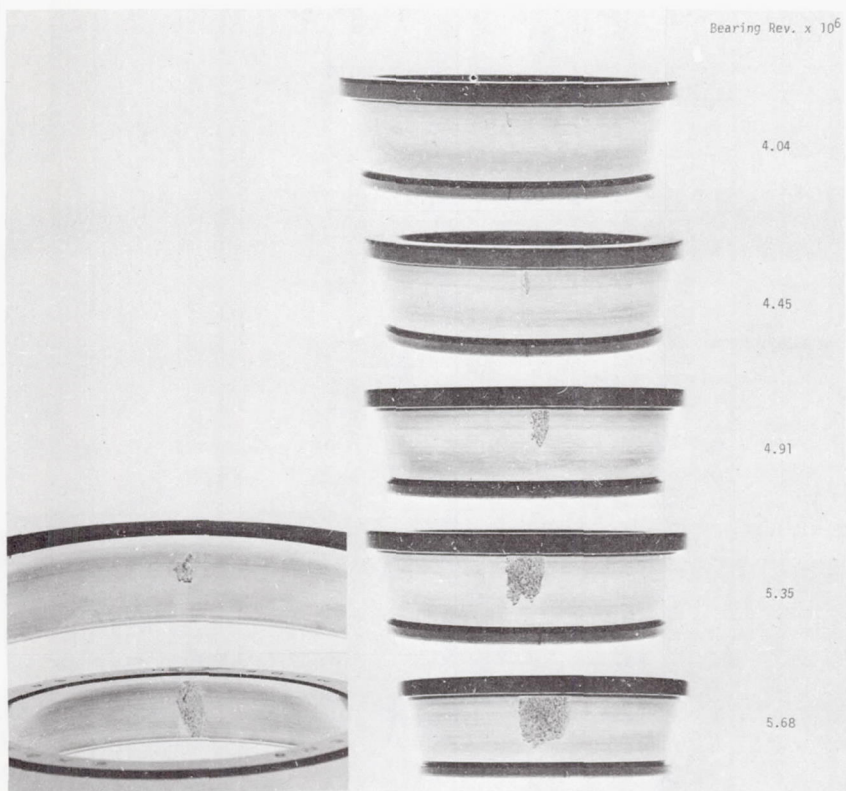


FIGURE 13.—Details of propagation studies with  $L_D$  and  $L_p$  in Mil-L-7808. Additional propagation beyond the 0.01-in.<sup>2</sup> laboratory criterion of failure is shown with the total number of cone revolutions beside each photograph.

#### HYDRAULIC PRESSURE PROPAGATION

When lubricant viscosity is low at the operating temperature and when the contact surface material has relatively high fracture toughness, a mode of propagation is seen that suggests a hydraulic-pressure-propagation (HPP) mechanism. Remarkably similar are the characteristics of surface-origin spalls formed in combined rolling and sliding, spalls formed under low EHD film thickness/surface roughness ratios in nominally pure rolling, and spalls formed by propagation of subsurface-inclusion-origin damage in a chemically aggressive lubricant of low viscosity at the operating temperature. Such spalls are compared in figure 14, and all show a characteristic arrowhead shape. On sectioning, many such spalls show a crack path that is unrelated to the known distribution of shear stress or any other calculable stresses except those generated by the lubricant trapped in the crack. Whether or not the actual mechanism involves

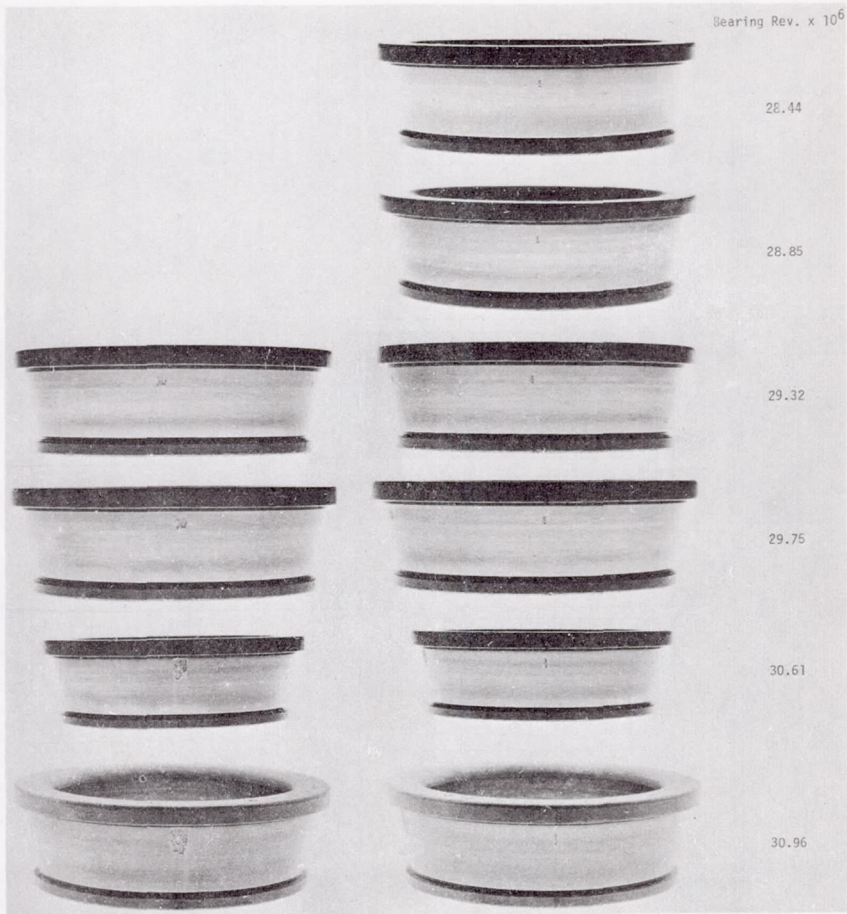


FIGURE 13.—Continued.

hydraulic pressure as the primary source of the stress operating at the crack tip, the propagation behaves as if it were at least governed by access of the lubricant to the perimeter of the propagating cracks beneath the surface.

Foord, Hingley, and Cameron (ref. 38) have recently attempted to test a hypothesis similar in some respects to the one originally proposed by Way (refs. 39 and 40). Both hypotheses utilize a propagation model in which a crack open to the surface is filled with the lubricant which is pressurized by the contact area crossing the mouth of the crack. One can visualize the moving contact progressively squeezing the fluid toward the edge of the subsurface crack, which is beyond the advancing contact stress field when the pressure pulse reaches it. Thus, without a compressive



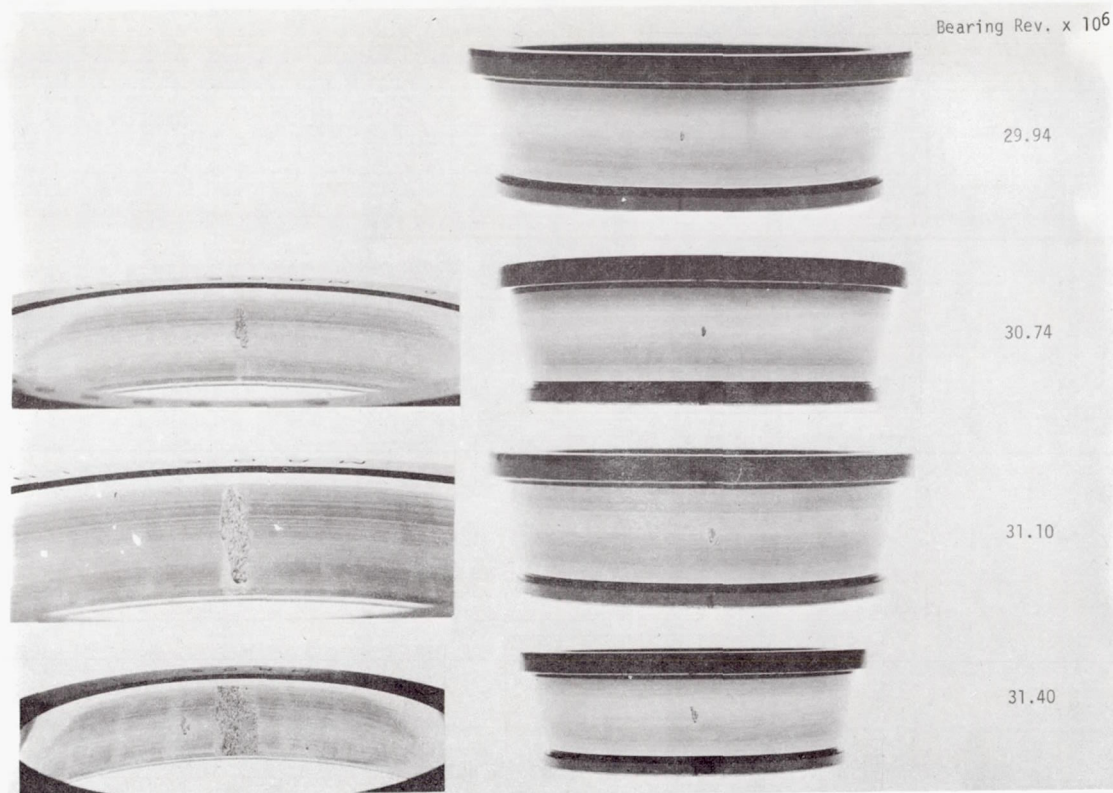


FIGURE 13.—Concluded.

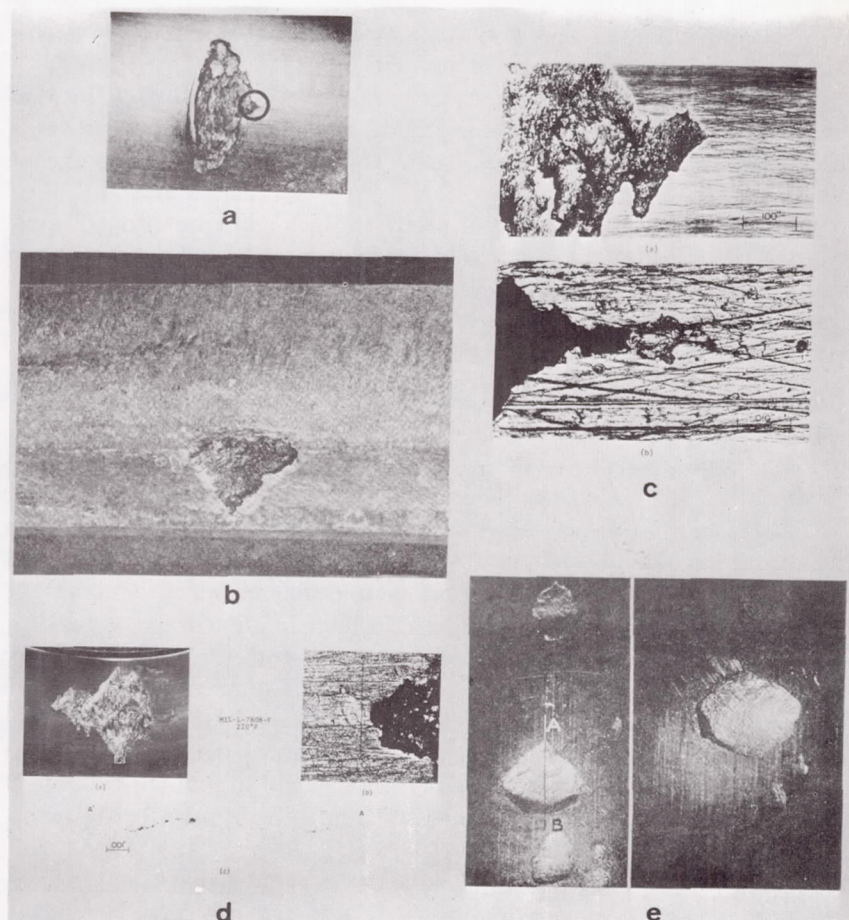


FIGURE 14.—Spalls having the appearance suggestive of hydraulic pressure propagation: (a) from Martin et al. (ref. 17), nominally pure rolling in a deep groove ball bearing; (b) from Davies and Day (ref. 25), rolling plus sliding in an aircraft gear; (c) from reference 35, PSO mode from a debris dent, nominally pure rolling in a tapered roller bearing; (d) from reference 35, inclusion origin plus HPP, nominally pure rolling in a tapered roller bearing; (e) from Way (ref. 39), 0.45% C gear steel in nominally pure rolling in a laboratory rig test.

stress to restrain it, the high pressure acts as a wedge exerting a tensile force on the sharp crack tip and causes a small fracture along the entire crack front with each stress cycle.

Observations that support this hypothesis are listed as follows:

(1) Evidence of a critical crack length for rapid propagation. The crack must extend beyond the compressive field of the contact stressed zone for the HPP mechanism to operate.

(2) The only cracks that propagate by HPP are those inclined downward in the rolling direction at a small angle.

(3) Relief of the pressure by branching cracks to the surface, into an undercut, by drilling a hole or removing the fluid from the system prevents the propagation or greatly retards it. (Mitsuda quotes the results of Way in ref. 15.)

(4) Low viscosity or low speed promotes HPP. The quantity of fluid entering the crack between stress cycles is governed by the atmospheric viscosity of the fluid and the access time. In soft steels under nominally pure rolling, a pitting endurance limit was observed at viscosities and speeds below that necessary for EHD film separation (ref. 38).

Alternative explanations are also possible. A Rebinder\* effect from polar additives, or impurities in the lubricant may reduce the elastic limit by promoting the exit of dislocations from the affected surface near the crack tip, or by reducing the interfacial energy between the fluid and the crack surface. Strongly adsorbing molecules are effective in reducing the fatigue life of mild steel above the fatigue limit in dry argon according to Duquette and Uhlig (ref. 41). They showed that oxygen in combination with water not only reduced fatigue life but eliminated the endurance limit in AISI 1015 steel. In aerated water or 3-percent NaCl in aerated aqueous solution, the bending fatigue cracks in their studies followed crystallographic planes in the ferrite, suggestive of a Rebinder effect. Galvin and Naylor (ref. 42) have shown a progressive decrease in bending fatigue life of steel tested in solutions of oleic acid in paraffin oil from 2 parts per million to 2 percent. Increases from 2 to 20 percent caused no further reduction in fatigue life.

Schatzberg and Felsen (ref. 37) have demonstrated that small quantities of water dissolved in a hydrocarbon lubricant can have significant effects on contact fatigue of hardened 52100 steel. Spitzig and Wei (ref. 43) found a 50-percent greater fatigue crack growth rate for high strength (300 ksi ultimate tensile strength) 0.45-percent carbon steel in humid argon or air compared to dry argon for fatigue notched specimens in uniaxial repeated tension tests. The effect was less for the same steel tempered at higher temperature and having higher fracture toughness.

Hydrogen embrittlement is widely prevalent, and steel at the carbon and hardness levels typical of contact surfaces is highly susceptible; so, a model based on corrosion at the crack tip with embrittlement of that region is attractive but difficult to verify. Scott, Loy, and Mills (ref. 22)

---

\* The effect of surface-active materials on the mechanical behavior of crystalline materials has come to be called the Rebinder effect after the author of extensive publications on the subject. Reductions in elastic limit and yield and fracture strengths have been reported in the presence of polar molecules or chemical effects other than corrosion.



have presented electron fractographic evidence of hydrogen embrittlement for 52100 type steel tested in a water-saturated lubricant.

Clearly some imaginative work is needed to clarify whether a hydraulic pressure mechanism of contact fatigue crack propagation does prevail under any practical circumstances. Since propagation is extremely rapid when it seems to happen and because its effect would not otherwise be expected in heavily loaded applications at low speed in high viscosity oils, the problem has real practical import. It is hoped that embrittlement, chemical effects, and mode of fracture will be disclosed with the help of such new tools as the scanning electron microscope (SEM). Figures 15 and 16 are examples of the type of information that can be obtained from the SEM and by optical microscopy for surface and subsurface origin modes of damage. Surface damage and subsurface contact fatigue spalls have been studied, and valuable results have been obtained from using replica methods by Scott et al. (ref. 44), Borgese (ref. 45), and Reichenbach et al. (ref. 46). As experience is gained in the correct interpretation of contact fatigue spalls, the convenience of direct examination in the SEM should offer great advantages over replica fractography in the transmission electron microscope.

#### SUBCASE FATIGUE DAMAGE

The mechanism of "case crushing" has been adequately described by Pederson and Rice (ref. 47) and Littmann and Widner (ref. 23). Its practical importance is mainly in gears and in heavily loaded contact fatigue rig testing of case-hardened materials. The term "case crushing" is unfortunate because it suggests a static overload effect in which yielding of the core leads to cracking of the case-hardened region. The evidence (refs. 23 and 47) suggests that this mode of contact fatigue is a special case of very deep subsurface-inclusion-origin crack initiation and growth.

Any contact surface that is case-hardened or has any gradient of strength vs depth below the contact surface should be designed to avoid this mode of damage. Figure 17 indicates how the cracks form deep beneath the surface where the ratio of applied shear stress to available shear strength is a maximum. Propagation can be extensive before surface-connected cracks allow spalling to occur. The resulting fragments tend to be large and can damage associated parts. This is of particular importance because destruction of the contact surface is very rapid, and there is little warning of trouble before spalling begins. Additional research is needed for parts such as rollers or multiple-contact rig tests in which more than one contact stress pattern affects the deep subsurface shear stress distribution. Superimposed bending or press-fit stresses would also be expected to modify the 0.55 (shear yield strength) criterion suggested by Pederson and Rice (ref. 47) and should also be investigated.

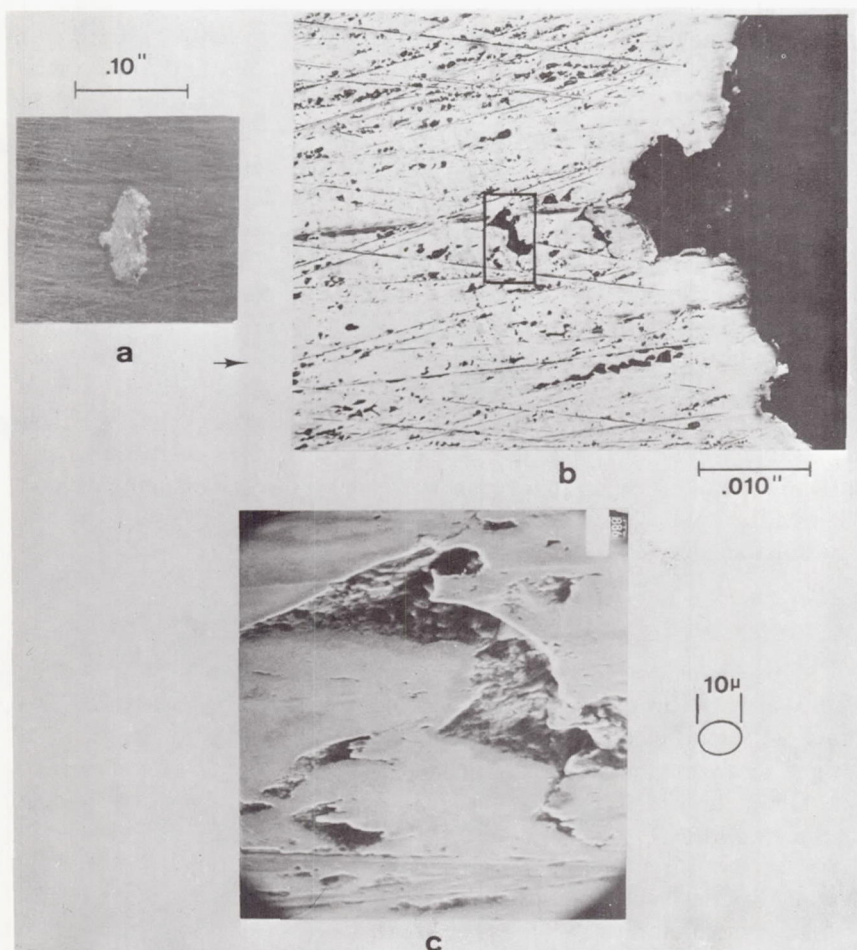


FIGURE 15.—PSO mode spall on a tapered roller bearing run in Mil-L-7808F. (a) Low magnification optical photograph of the spall, (b) higher magnification optical photomicrograph of spall origin, (c) scanning electron micrograph of the area indicated in (b). The image is foreshortened as shown by the  $\mu$  ellipse because of the  $45^\circ$  orientation of the sample in the SEM instrument.

#### SURFACE-ORIGIN DAMAGE

The mechanism of surface-origin contact fatigue damage under nominally pure rolling-contact conditions is understood in terms of the variety of surface flaws that can serve as the prime cause of local stress intensification leading to surface cracking. Debris dents, handling nicks, grinding furrows, or other flaws of surface texture, surface inclusions, corrosion pits, and peeling damage act as points of stress concentration from which



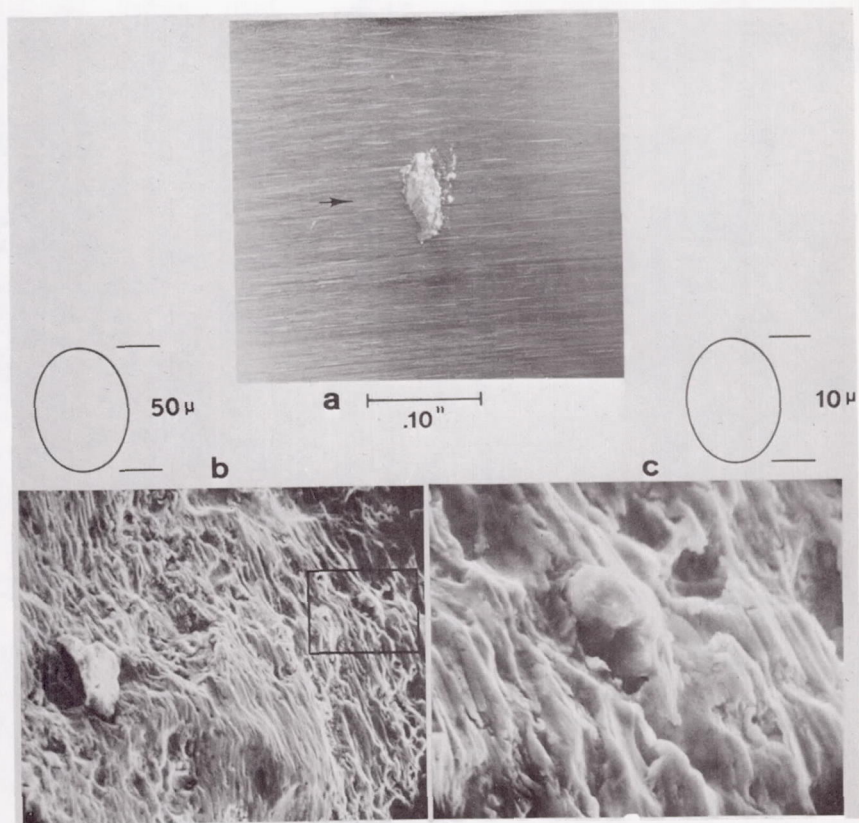


FIGURE 16.—Subsurface inclusion origin spall on a tapered roller bearing run on Mil-L-7808F. (a) Low magnification optical photograph, (b)  $600\times$  scanning electron micrograph within the spall, (c)  $3000\times$  SEM micrograph of the area indicated in (b). An inclusion particle can be seen and a hole where another particle had been nearby.

a rapidly propagating contact fatigue crack can originate. All of these have been described in published literature (refs. 3, 17, and 28), and examples of each are shown in figure 18.

In addition to a surface flaw that produces locally intensified contact stress, some additional factors are usually necessary for point-surface-origin modes of damage to become prevalent. Gross sliding in combination with rolling contact, as in gears, nearly always produces an arrowhead shaped point-surface-origin type of spall (refs. 16 and 48). Figure 19, according to Niemann et al. (ref. 48), shows the progressive formation of the typical PSO mode damage on gear teeth. The tensile stress ahead of the contact area on the surface which is in negative sliding is believed to open the crack, admitting lubricant and promoting surface crack



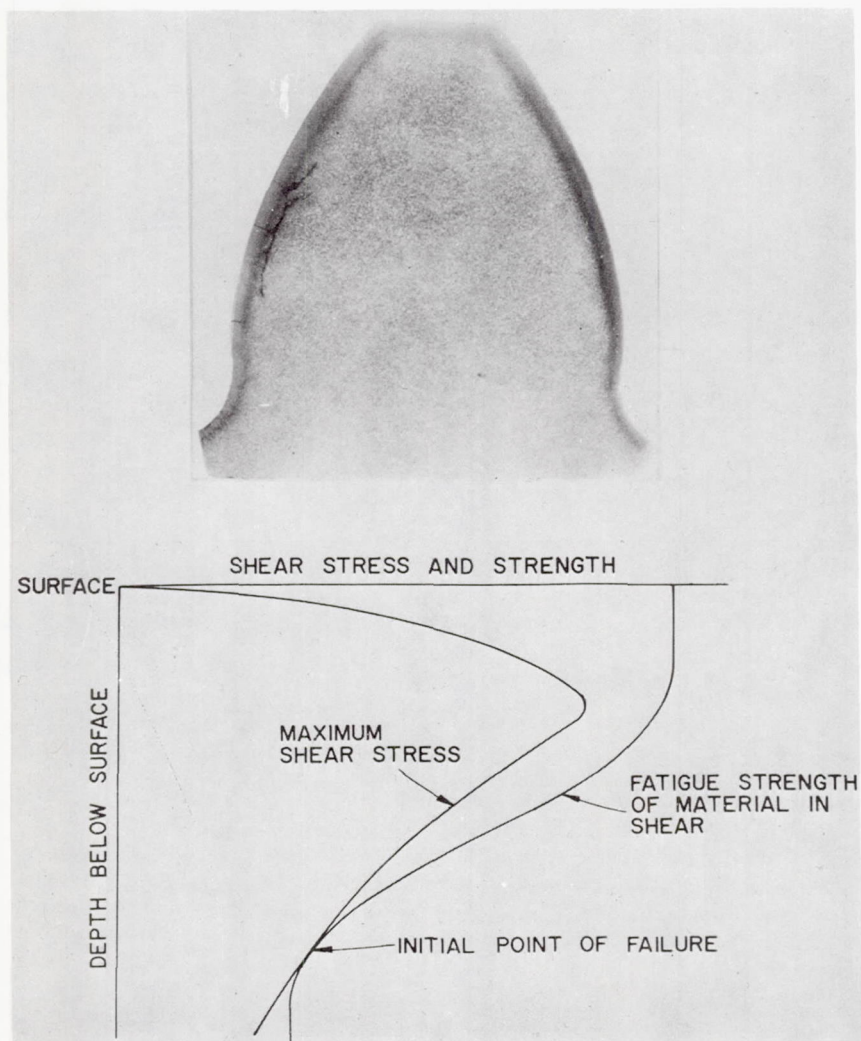


FIGURE 17.—Subcase fatigue, called case crushing by Pederson and Rice (ref. 47).

This illustration from their paper shows the early stages of fatigue cracking in the lower carbon portion of the carburized case. The diagram below indicates how the stress and strength gradients combine to cause the weakest condition at the subcase location where fatigue cracking begins.

propagation. This leads one to speculate whether sliding or at least large tangential forces must be present for surface cracks to form. Stress analysis makes it clear that the maximum shear stress range is at the surface when tangential forces display an effective friction coefficient 0.1 or greater (ref. 4). Numerous studies of surface-origin damage have

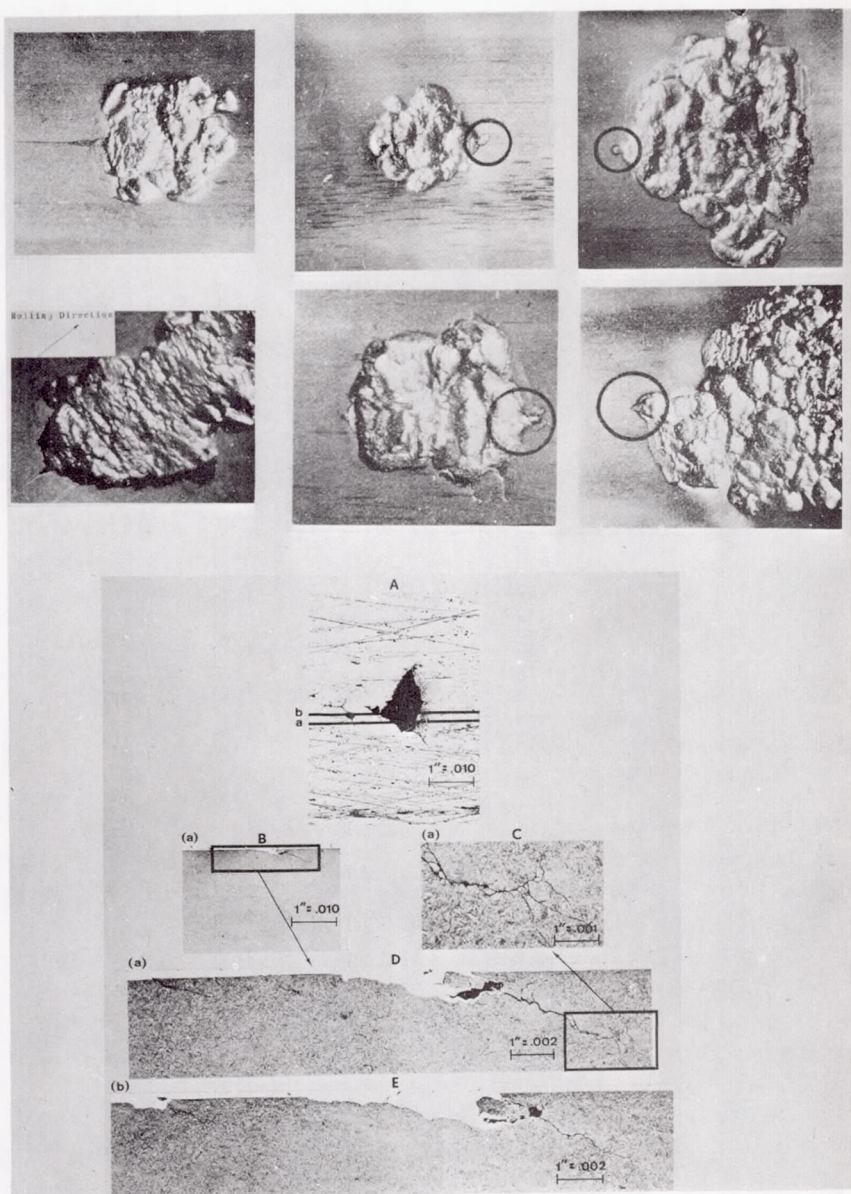


FIGURE 18.—PSO modes of contact fatigue damage formed under nominally pure rolling. (a) Origin of spalls at debris dents, pits, and grinding furrows in ball bearings are from Martin et al. (ref. 17); (b) the origin was at one of many general surface damage mode cracks on this tapered roller bearing cone which ran in Mil-L-7808 at 170° F with  $\frac{h_G}{\sigma_1} = 0.44$ . Cross-sections show the nature of crack propagation. See also figures 10C and 11b for PSO type spalls from inclusion origins.

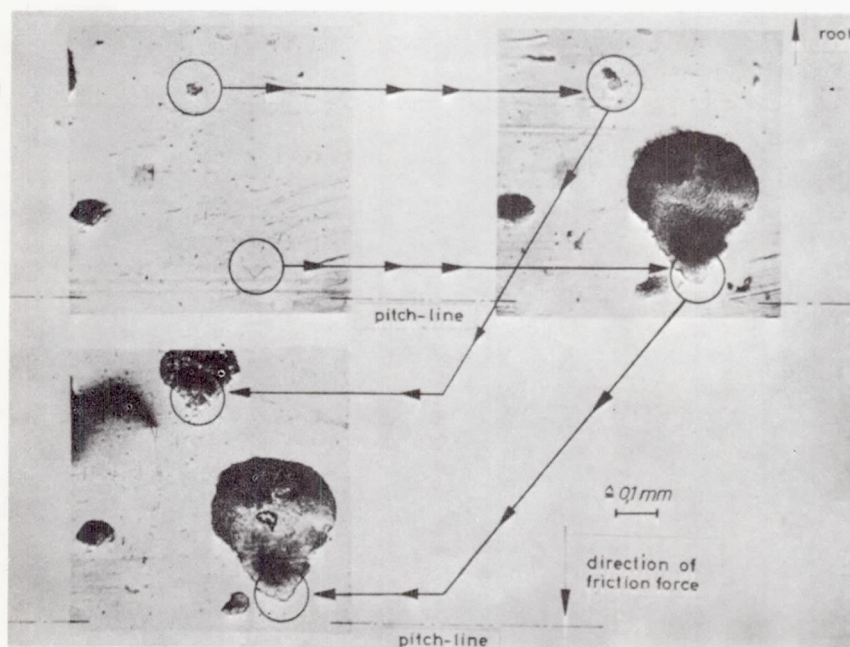


FIGURE 19.—Progressive development of PSO mode contact fatigue damage under combined rolling and sliding on a gear tooth according to Niemann et al. (ref. 48). See also figure 14b.

explored the effects of surface texture, lubricant composition, EHD film thickness, and other variables under a variety of rolling plus sliding conditions. Unfortunately few of them have studied the mode of fatigue damage in high-carbon, high-hardness steel for a range of sliding velocities including zero sliding. It is clear that the first cracks are always formed at the surface under combined rolling and sliding (refs. 17, 18, 23, 46, and 48); but the least magnitude of sliding or tangential force necessary to cause surface-origin contact-fatigue cracks is not known. An EHD film thickness sufficient to prevent asperity interactions can completely suppress surface cracking (refs. 38 and 49); but as Kelley\* has commented, most practical situations where surface fatigue is critical operate in the mixed EHD film-asperity contact region. Operation in this range is certainly one of the factors which promotes surface-origin modes of damage.

Freedom from competing subsurface stress raiser as provided by CV remelted or other premium-quality steels will virtually eliminate inclusion

\* See Kelley's lecture on Lubrication of Concentrated Contacts—The Practical Problem.



origin modes of contact fatigue, and surface damage modes will then become prevalent (refs. 3, 17, and 23). Low lubricant viscosity, low EHD film thickness/composite surface roughness ratio, elevated temperature, and chemically aggressive base stock, additives or contaminants in the lubricant or atmosphere also promote surface cracking. Reichenbach et al. (ref. 46) demonstrated that surface cracks form and propagate as a function of the lubricant composition under conditions of combined rolling and sliding on hardened 52100 steel balls. Although no data on EHD film thickness/roughness ratios were given, the disappearance of original surface texture indicated mixed EHD-asperity contact conditions. Etching of contact tracks revealed microcracking at undissolved carbide-matrix interfaces and linking of cracks from adjacent carbides. They also showed that 0.5-percent sulfur in type 1010 petroleum oil caused sufficient chemical wear to remove surface microcracks and reduce contact stresses by providing a conforming track, a factor to be considered carefully in ball type test rigs.

The mechanism of surface-origin crack propagation is not well understood. When conditions are such as to promote surface cracking, the number of cracks is often very great. The growth of such cracks appears to be slow even though they appear after a relatively small percentage of the number of stress cycles to failure (ref. 46). Figures 15 and 18b show the large number of surface cracks near the origin of PSO type spalls on case-hardened 4620, modified steel. The nature of the crack propagation seen in a cross-section of the spall in figure 18b is identical with that observed when a subsurface-inclusion-origin spall propagates by the HPP mechanism (see fig. 10). It appears that the critical crack length for rapid propagation may be related to the contact length in the rolling direction, but no quantitative evidence is available to substantiate this view. At present, the mechanism of PSO mode crack growth appears as if it were by hydraulic pressure propagation, and the comments made previously apply here as well.

#### GENERAL SURFACE DISTRESS

Under some conditions of operation a general propagation of surface cracking is observed that is different from the PSO mode and yet is not the same as peeling. The propagation is deeper than peeling and appears to be a general occurrence of PSO type damage. Since the appearance of the damage at failure is different from PSO or peeling and it is likely that a different combination of circumstances promotes its occurrence, it is now considered as a separate damage classification. The PSO mode is probably a localized occurrence of the general surface damage mode due to a singular surface stress concentration which causes more rapid propagation to the critical size for HPP. When none of the cracks gets ahead of the others, adjacent cracks intersect before reaching the critical size

for HPP, thus providing leakage paths which prevent HPP. Fatigue damage occurs then by a progressive general propagation of intersecting surface-origin spalls of the type illustrated in figure 20.

A systematic appraisal reveals that the major variables of importance for control of surface damage are the following:

#### Surface Microtopography

In a surface free of flaws, the surface texture becomes increasingly critical as the EHD film thickness approaches the dimensions of the

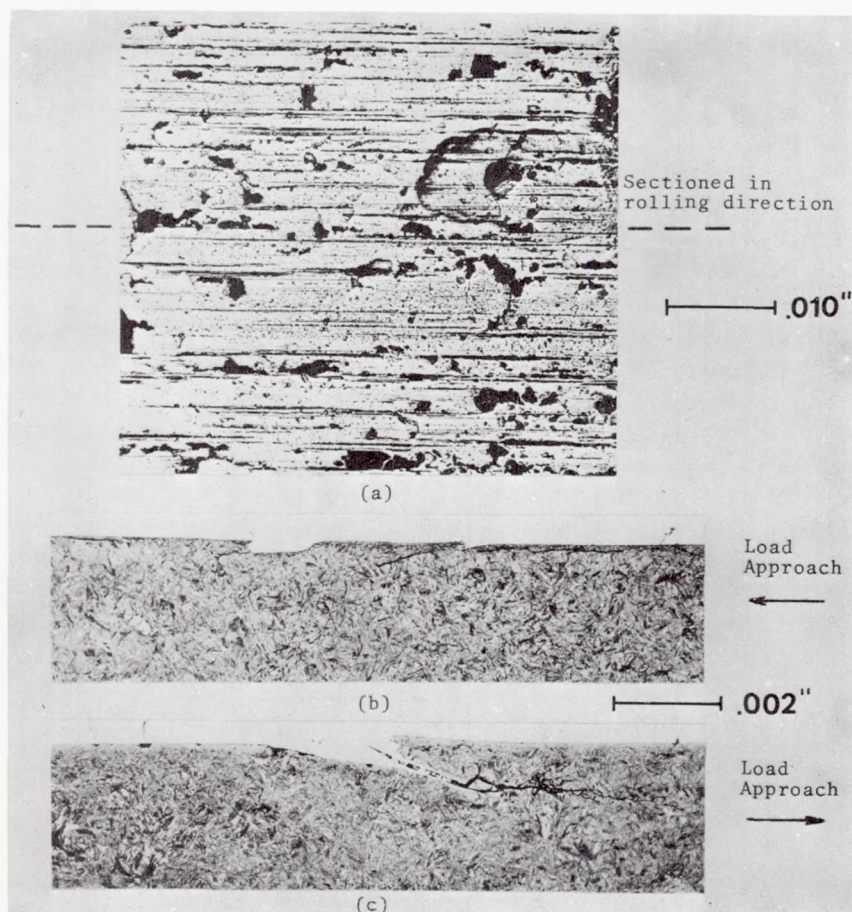


FIGURE 20.—General surface damage on a ground cone from one of the tests described in table 6. Cross-sections reveal a variety of crack depths in contrast with that seen when "peeling" occurs. The deeper crack in (c) may be a very early stage of PSO mode or HPP type propagation of the type seen in figures 15a and 18b.



surface roughness. The asperities in the surface microtopography then become the stress raisers at which surface cracks nucleate. Elegant studies of asperity dimensions, distributions, and interactions coupled with equally elegant studies of EHD film dimensions and related stress distributions have provided valuable insight and made possible an engineering approach to the control of surface-origin modes of damage related to surface texture. While evidence is strong that the ratio of computed EHD film thickness to composite roughness can be a useful index of the tendency to surface distress and surface-origin modes of contact fatigue, the fatigue life is not simply related to such a parameter except for a given contact surface material, contact geometry, surface texture, lubricant, and atmosphere composition. Furthermore, such mechanical effects as tangential forces, sliding, microslip, or cavitation in the lubricant can also play significant roles.

Indeed, until we have more adequate tools for quantitative characterization of surface microtopography, the rate of progress will be slow. What is needed is a means for computing the true contact stress distribution for surfaces in rolling and/or sliding contact in the range of EHD film thickness where the load is shared between asperity contacts and the EHD film. Tallian (ref. 13) has outlined this aspect of the problem and suggests that the radius of curvature or slopes of the asperities as well as their height distributions on the contact surfaces are important. Williamson (ref. 50) has demonstrated that only the top 15 percent of the asperity summits touch before bulk plastic flow occurs in the contacting bodies. Elastohydrodynamic pressures and the random relative positioning of surfaces in rolling contact may cause more of the asperities to interact, but it seems certain that the upper extremes of the surface microtopography have major influence on the surface stress distribution. Yet the valleys affect the local availability of lubricant, and the total profile affects the microscopic stress distribution in the surficial region. Table 6 compares the results of some bearing life tests which illustrate that the type of finish exerts its own effect in addition to that reflected in the EHD film thickness/composite roughness ratio when the operating conditions promote general surface damage.

#### **Elastohydrodynamic Film Thickness**

Within the last decade the documentation of EHD film thickness effects on contact fatigue have been so numerous and convincing no attempt will be made here to add to them, especially in view of the comprehensive treatment of this subject by Dr. Dowson.\* Microelastohydrodynamic phenomena are more elusive, but evidence is growing that some of the chemical effects of lubricants may be manifested in this area.

---

\* See Dr. Dowson's lecture on Elastohydrodynamic Lubrication.



TABLE 6.—*Effect of Surface Texture on Contact Fatigue Life*

	Test No. 3	Test No. 4	Test No. 5	Test No. 6
No. tested	18	30	30	30
No. failed	16	29	30	30
Cone texture	Honed	Ground	Ground	Honed
Cone roughness—Range, $\mu\text{in. AA}$	4.0–7.0	11.0–24.0	11.0–24.0	4.0–7.0
Ave., $\mu\text{in. AA}$	5.5	17.3	17.3	5.5
Roller texture	Ground	Ground	Ground	Ground
Roller roughness—Range, $\mu\text{in. AA}$	3.5–5.0	3.5–5.0	3.5–5.0	4.0–10.0
Ave., $\mu\text{in. AA}$	4.0	4.0	4.0	7.0
$h_G$ EHD film thickness (Grubin), $\mu\text{in.}$	4.02	4.02	9.02	2.16
$\sigma_1$ , composite roughness, $\mu\text{in. AA}$	9.5	21.3	21.3	12.5
$\Delta G = h_G/\sigma_1$	0.42	0.19	0.42	0.17
Temperature (outlet oil), $^{\circ}\text{F}$	170	170	125	160
Max. compressive stress, ksi	325	325	325	325
Speed, rpm	2700	2700	2700	900
Weibull life estimates:				
$L_{10}$ rev. $\times 10^6$	4.8	2.1	5.7	1.3
$L_{50}$ rev. $\times 10^6$	38.9	11.8	17.2	3.7
Weibull slope, $m$	0.91	1.09	1.70	1.75
Correlation coefficient, $R$	0.99	0.95	0.98	0.95
Predominant damage mode	13 cones PSO + GSC 2 rollers PSO 2 rollers GSC	20 cones PSO + GSC 13 rollers PSO + GSC	15 cones PSO 18 rollers PSO	30 cones PSO All rollers OK

Tapered roller bearing, 1.02-in. mean cone, O.D. 0.24-in. mean roller O.D., 4619 (modified) case carburized and hardened to 60 Rc before finishing.

Thrust loaded in 3-in. O.D. Timken bearing life test machines.

SAE 20 petroleum oil lubricant flow at 1 pint per minute per bearing.

Inlet oil temperature was controlled to the same as outlet oil temperature after the load-up period.

### Temperature

Since temperatures vary from point to point within the contact stress affected region and related thermal expansion gradients impose stresses that modify those arising from contact loads and elastohydrodynamics, it is certain that temperature distributions within the lubricant film and contact surficial regions are important. Flash temperatures in rolling plus sliding are discussed by Professor Blok in this symposium, but principally as related to lubrication failure. Kelley (ref. 51) and others have examined the cyclic stresses caused by thermal gradients in the contact surfaces as they may contribute to contact fatigue. Such stresses are insignificant except for appreciable sliding velocities and become more damaging as the degree of asperity contact is increased by low EHD film thickness or rough surface texture.

Temperature is also important as it affects the rate of all relevant chemical reactions between the lubricant and the contact surface materials. Crack initiation and crack propagation are both affected. Surface films that may aid in preventing crack initiation are formed more readily at elevated temperatures so that beneficial effects can be operating with respect to surface crack initiation; at the same time the same mechanism could promote crack propagation by preventing rewelding at the sub-surface crack tip.

High temperatures promote degradation of the lubricant by oxidation in air, forming acids that have been shown to decrease contact fatigue strength (ref. 52). Additives are depleted or altered by chemical reactions with the atmosphere or other materials in the system. If the reaction product is less detrimental to contact fatigue life than the additive, unexpected beneficial effects may be present. All of these possibilities must be considered in the evaluation of the effects of temperature.

Finally, the effect of temperature on lubricant viscosity is of major importance. Bulk viscosity at the rolling-contact inlet is governed by lubricant temperature at that point, not always easily determined but strongly affecting EHD film dimensions. Within the macro-contact area the temperature distribution affects local viscosity at asperity interactions, and side leakage effects at surface flaws and at the periphery of the contact area.

### Lubricant Composition

Base stock composition governs viscosity variations with temperature and pressure and thus modifies contact pressures by elastohydrodynamics and temperature rise in the contact region (ref. 53). Chemical activity has already been discussed under temperature effects but is controlled by selection of base stock and additive package composition (refs. 52 and 54). Oxidation, thermal degradation, or contamination change the composition of the lubricant during service or testing and thus affect

surface-fatigue damage. The difficulties of attempting to compare the relative influence of lubricants by means of contact fatigue have been discussed by Koved (ref. 55).

Since other papers in this symposium will deal with such effects in detail, my concern here is primarily for the mechanism by which lubricant composition affects surface-origin modes of damage. This has been discussed previously under subsurface-inclusion-origin damage propagation but is even more important for surface-origin modes, since crack initiation as well as propagation are under the influence of the lubricant from the first stress cycle onward.

Galvin and Naylor (ref. 42) have demonstrated that the fluids they tested could be classified into two groups, Class I fluids which affected finite fatigue life but did not affect the bending fatigue limit (0.15 percent C, 2.9 percent Ni, 0.3 percent Cr annealed to 140 VPN) and Class II fluids which affected both finite life and fatigue limit. Using the same fluids with high-purity aluminum, phosphor bronze, gold, and platinum on electropolished and ground surfaces tested in bending fatigue led to the conclusion that the major factor affecting crack propagation rate for a reactive metal was the rate at which the fluid could penetrate the crack. In both crack initiation and crack propagation, the fluid must be capable of reacting chemically with the metal or the oxide layer on the surface to exert any influence. The mechanism by which crack initiation was affected by oleic acid could not be related to chemical attack since temperature exerted no effect on fatigue life until oxidation of the oil at 190° C produced measurable corrosion fatigue. Yet exclusion of contact by gold-plating prevented the oleic acid from affecting the fatigue limit. As little as 0.0002 percent of oleic acid in mineral white oil produced a significant decrease in fatigue life of the steel used in their tests.

The importance of polar components and of subtle corrosion effects in fatigue crack initiation processes suggests a Rebinder effect that lowers the stress for dislocations to emerge from the surface. The data of Duquette and Uhlig (ref. 41) for a 1018 steel also support a Rebinder effect on bending fatigue crack initiation when corrosion effects are absent. On the other hand, they also demonstrated by cathodic protection that corrosion was primarily responsible for the elimination of the fatigue limit when tests were run in aerated water or 3 percent NaCl solution.

#### Atmosphere Composition

The primary effect of common gaseous atmosphere composition is whether it promotes oxidation of the lubricant or additive package. Other effects are primarily due to oxygen and water because they may be physically adsorbed or react chemically with the contact surface and propagating crack.



### Load and Contact Geometry

The level and distribution of contact stresses are primarily governed by load and contact geometry, modified by surface texture and EHD effects. The level of contact stress is important because so many accelerated tests are run in the range of full plastic deformation. There is no assurance that the relative ranking of materials or lubricants as they affect contact fatigue performance can be extrapolated over wide ranges of load or contact geometry unless some assurance is found that the mode of fatigue damage is unchanged. It is possible that general-surface-initiated damage makes a gradual transition to peeling as load is reduced. The relative stress concentration effects of surface flaws and subsurface inclusions probably become greater at lower levels of contact stress.

### GEOMETRIC STRESS CONCENTRATION

The simplest form of localized contact fatigue damage obviously related to macroscopic contact stress geometry is seen in figure 21 according to Lucas (ref. 56) and Simpson (ref. 57). End of contact damage in line-contact types of bearings is seen most often under conditions of misalignment, but examples of GSC damage have been seen in well-aligned tapered roller bearings and rig test specimens (refs. 3, 19, 24, and 58). In its earliest stages, cracking is usually seen at the surface, probably

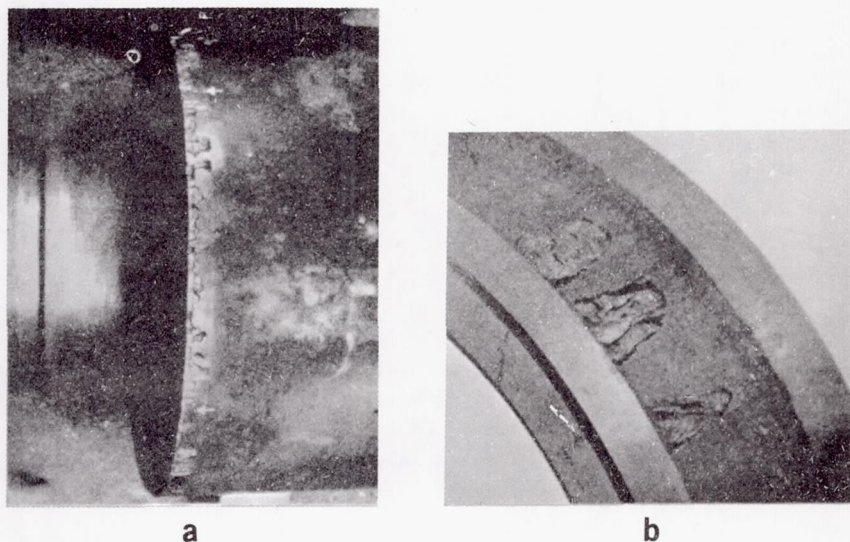


FIGURE 21.—(a) Geometric stress concentration mode of contact fatigue damage on a steel mill backup roll according to Lucas (ref. 56). (b) GSC mode spalling on a cylindrical roller bearing run under excessive deflection or misalignment, according to Simpson (ref. 57).

because of tangential forces as well as the radius of curvature at the end of the contact. Moyer and Neifert (ref. 59) have described a simple method for computing the stress concentration effect at the end of line contact, and McKelvey and Moyer (ref. 58) illustrated the relation of peak compressive stress to contact fatigue life for toroidal rollers between cylindrical disks.

The crowning of rollers in line-contact bearings, of gears, or of steel mill rolls is the simplest means used to avoid GSC modes of damage. It is recognized that modifications of internal bearing or gear geometry to overcome GSC effects caused by end of contact geometry coupled with deflections or misalignment are effective over a limited range of load for which they are designed.

The coincidence of subsurface nonmetallic inclusion-spall origin near the end of contact can be mistaken for GSC. If no earlier stages of pitting are seen along the end of contact region elsewhere on the bearing, the spall is probably of inclusion origin. Examples of true GSC and inclusion origin at the end of contact are illustrated in figure 22.

Sometimes the occurrence of peeling near the end of contact will destroy the original surface and cause a more severe stress concentration than was originally present. The edge of a peeled area may create a GSC mode of damage well away from the end of the original contact path as seen in figure 23.

#### PEELING

The shallow contact fatigue damage which propagates by progressive superficial pitting is sometimes called "frosting" because of the dull texture of the affected areas. The remaining surface often shows a burished appearance, called "glazing." Peeling is characterized by the formation of cracks that originate at a very shallow angle to the surface and propagate parallel to the surface at depths that are very small compared to the depth of maximum shear or orthogonal shear stress range. It is most easily produced by rolling with a small percentage of sliding when the composite surface roughness is greater than the EHD film thickness. Gross sliding is not required to cause peeling; but microslip or tangential force is probably involved, at least at individual asperity contacts. Crack depths are in the range of 0.0001 to 0.001 inch and on the rougher surface cracking is confined to the asperities if there is a significant difference in surface roughness between the contact surfaces. The smoother surface usually suffers the most extensive damage. Examples of peeling from Simpson (ref. 57) are illustrated in figure 24.

Cross-sections of early and advanced stages of peeling in figure 25 suggest that the reversed plastic strain associated with each asperity contact of the type described by Shaw (ref. 60) is the probable cause of the uniformly shallow cracking we call peeling. The depth of peeling



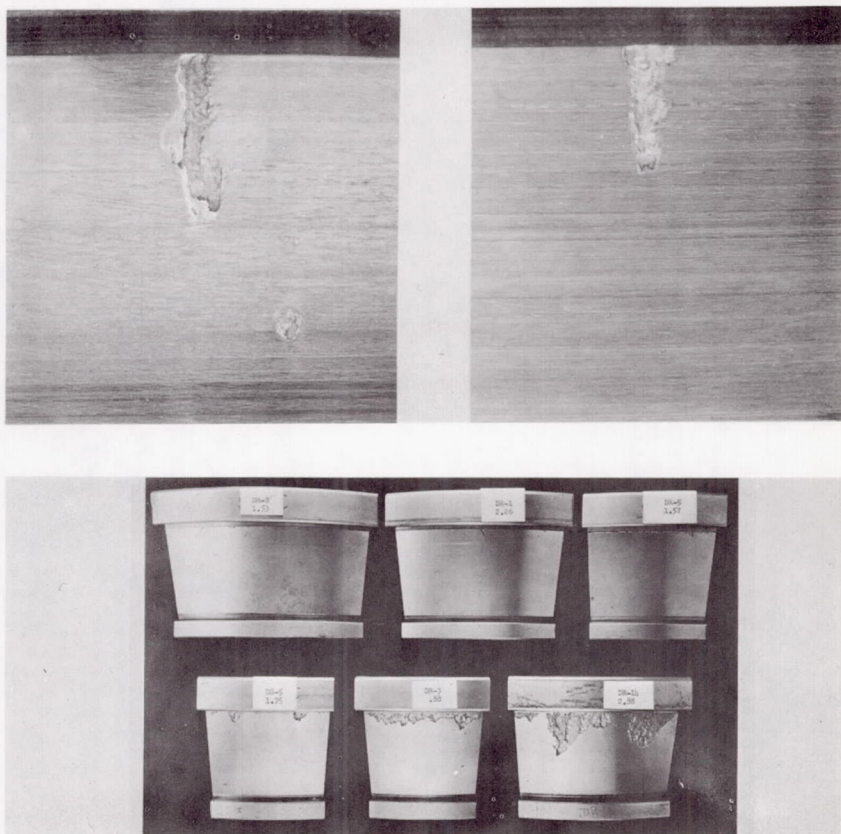


FIGURE 22.—A comparison of inclusion origin spalls near the end of contact (above) with true geometric stress concentration spalls (below) on tapered roller bearings (ref. 23).



FIGURE 23.—Geometric stress concentration spalling at the edge of a peeled area on a tapered roller bearing cup, according to Wren and Moyer (ref. 19).



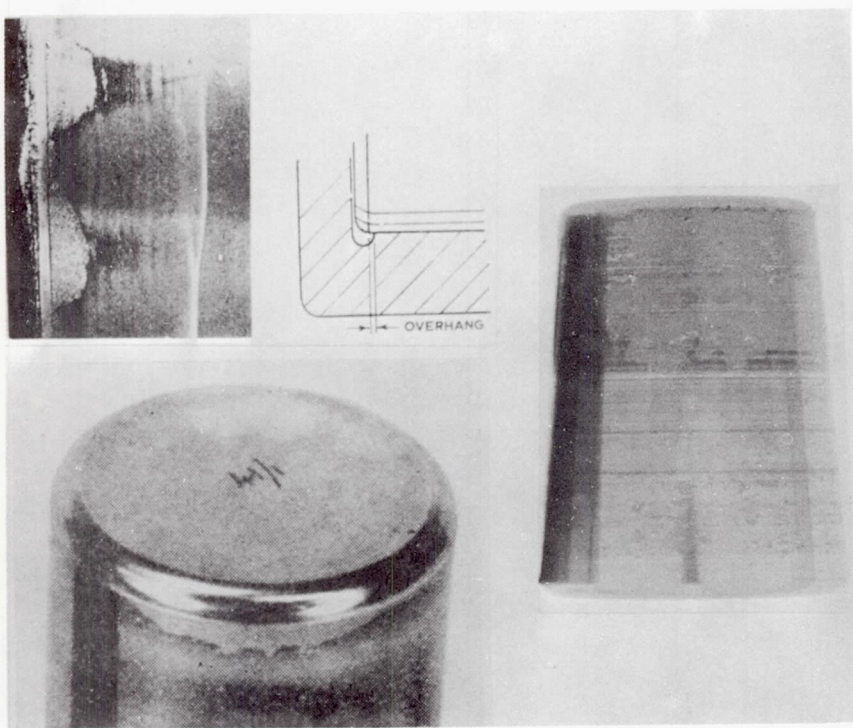


FIGURE 24.—Peeling damage on cylindrical bearing rollers according to Simpson (ref. 57), left, and on a tapered roller (ref. 23), right.

seems to be less for lower values of composite roughness, although no quantitative information is available to substantiate this observation.

Localized thinning of the EHD film can lead to small regions of peeling at furrows in the surface (refs. 17 and 23). Side leakage of lubricant at the edges of the contact area seen by Gohar and Cameron (ref. 61) leads to peeling of the type in figure 26 only if the composite roughness is critical for the prevalent lubrication conditions. For more favorable surface conditions, a localized smooth wear can occur in the region of local EHD film leakage with no fatigue damage. Lubricant additives may exert an influence on the tendency to peeling or smooth wear, all other things equal.

It seems most likely that peeling occurs when a relatively large percentage of the surface asperities in the surface texture are sharing the load with the EHD film, but at a low level of nominal contact stress so that plastic deformation is confined to the immediate subsurface region under each asperity contact. Dark etching effects resulting from such near-

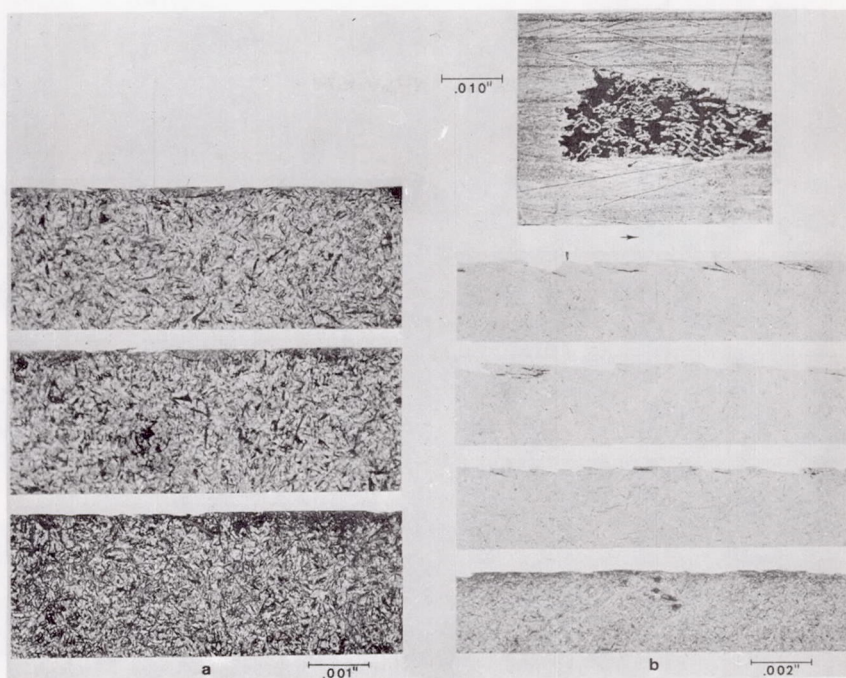


FIGURE 25.—(a) Radial cross-sections of early stages of peeling damage in a tapered roller bearing cone. Sections are perpendicular to the rolling direction. (b) Advanced peeling damage on a tapered roller, above, with cross-sections of the same roller, below. The plane of these sections is parallel to the rolling directions indicated by the arrow.

surface plastic deformation can be seen in figure 25 and seem to correspond in depth to the range of peeling crack propagation.

#### SECTION FRACTURE

When cross-sections of bearings or gears are subjected to tensile or bending stresses, there is a tendency for contact fatigue cracks to propagate through the section. Excessive press fit will likewise promote section fracture. Such fractures are seen more often in through-hardened sections than in case-hardened sections because an intermediate region of higher fracture toughness in the latter retards crack propagation. Moore et al. (ref. 62) developed a stress analysis that was helpful in preventing section fracture when it was encountered in flange-mounted ball bearings of M-50 steel.

The mechanism of section fracture, in the author's opinion, is simply the propagation of cracks from the contact fatigue region by cyclic tensile stresses due to press fit or flexing of the section. Since one of the planes



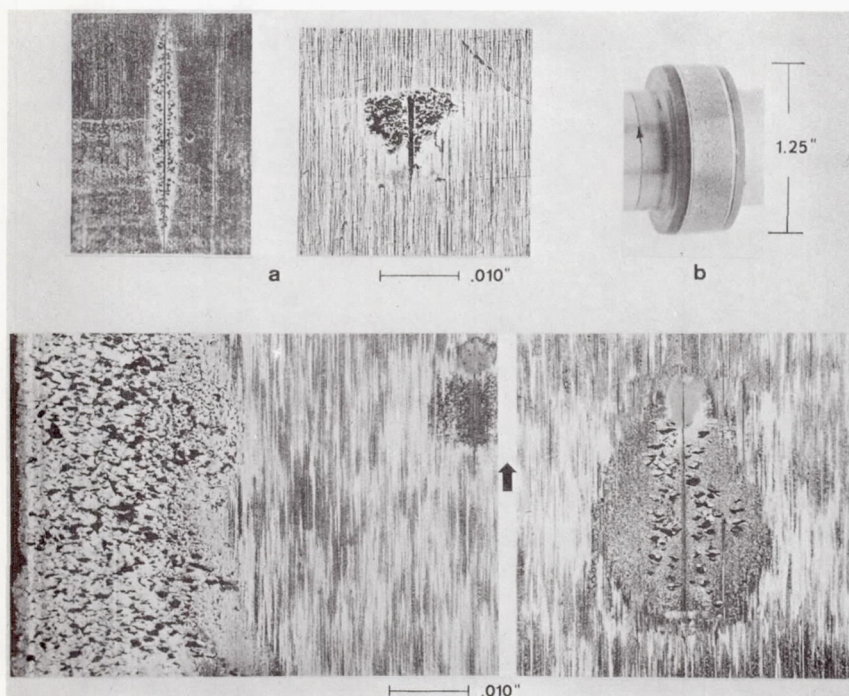


FIGURE 26.—Peeling damage due to localized EHD film leakage. (a) Peeling at grinding furrows, according to Martin et al. (ref. 17) and from reference 23; (b) peeling at the end of contact and at grinding furrows on a rig test specimen run in SAE 5 petroleum oil with  $h_0 = 2.8 \mu\text{in.}$  at 620 rpm. General appearance after  $186.1 \times 10^6$  stress cycles at 350 ksi max. compressive stress.  $1\frac{1}{4}$ -in. diameter specimen was run against a 3-in. diameter drive roller having 100-in. crown to provide a truncated elliptical contact. Microscopic appearance of damage is shown below.

of maximum orthogonal shear stress range is oriented in a radial direction in a bearing, the section fracture can occur before any visible propagation of contact fatigue damage on the raceway (ref. 19). More often, the section fracture is a secondary mode of crack propagation after subsurface- or surface-origin contact fatigue damage to the raceway (ref. 62).

Section fracture can be minimized by the use of higher fracture toughness core materials, as in case-carburized bearings and gears. Control of tension or bending stresses on sections subject to contact fatigue are the most positive means to prevent this mode of damage.

#### ROLE OF ALTERATIONS OF MICROSTRUCTURE IN CONTACT FATIGUE

The metallographic studies of contact fatigue have failed thus far to establish whether the dark etching alterations (DEA) or light etching alterations (LEA) play any significant role in the nucleation of contact



fatigue cracks. The LEA effects have been discussed earlier in this paper. The studies of Schlicht (ref. 29), Widner (ref. 23), Martin et al. (ref. 28), Bush et al. (ref. 33), and Gentile et al. (ref. 34), have shown that there is a threshold stress for the general appearance of LEA. Its distribution is related to  $\tau_{45}$ ; the maximum unidirectional shear stress and the frequency of LEA, whether general or associated with nonmetallic inclusions, increases with the contact stress level and the number of stress cycles. Carter (ref. 30), Bush et al. (ref. 33), Gentile et al. (ref. 34), Jones (ref. 63), and others have made similar observations for the DEA in 52100 steel. Temperature is influential on the stress level at which the DEA forms. It seems that the DEA appears generally in carburized bearing and gear steels much in the same manner as the general occurrence of LEA in 52100 and other through-hardening bearing steels. The higher retained austenite level normally seen in the carburizing steels and the undissolved carbides normally seen in 52100 type steels may contribute to this difference in microstructural response to cyclic contact stress.

The transmission electron microscope studies of O'Brien et al. (refs. 12 and 32) suggest that the LEA and DEA are similar in several bearing steels and that both effects are the result of localized regions of BCC ferrite having a high density of dislocations arranged in a cellular structure with an absence of visible temper carbides. Cell sizes observed were 0.05 to  $0.1\mu$ , much smaller than seen in cellular dislocation arrays resulting from plastic deformation in other materials. Their results clearly showed that both of the altered microstructures are the result of plastic deformation. In case-carburized 4620, modified steel the DEA consisted of microband subgrains that spread from one martensite plate to another, aligned in a common direction.

The cellular dislocation network seen by O'Brien et al. (refs. 12 and 32) in carburized 3310 steel within the butterfly alteration at inclusions appears very similar to that shown by Martin et al. (ref. 28) in 52100 steel within the LEA at deformation bands found generally in the sub-surface region after cyclic stressing. The cellular dislocation structure displays a high solubility of carbon, believed to be associated with an interaction of the stress field of the individual dislocations with the interstitial carbon atoms. Carbon mobility is greatly increased during cyclic stressing in heavily loaded rolling contact, evidenced by the growth of lenticular carbides (ref. 28).

All of the above observations are clearly related to the magnitude and distribution of shear stresses in concentrated contacts, yet these stress-induced alterations of microstructure do not appear to be fatigue damage in the sense that they precede crack formation or are an essential step in contact fatigue crack initiation. The fact that the threshold stress for appearance of LEA correlates with the elastic limit and the location of LEA corresponds with the maximum unidirectional shear stress distri-

bution indicates that it is the result of localized plastic deformation in regions favorably oriented for slip or dislocation motion. The sensitivity of DEA to temperature as well as cyclic shear stress and a better correlation of its distribution with the maximum orthogonal shear stress range make it more difficult to understand.

Both the LEA and DEA appear most useful as a means to identify locations of stress concentration in hardened steel subject to contact fatigue. Through correlations with contact stresses, for a given material and pretest microstructure, these microstructural alterations become a convenient internal strain gage that can be used to study stress distribution in *concentrated contact*. Perhaps they will be useful in studying the detailed distribution of contact stresses in the mixed regime of thin EHD film and partial asperity contact.

#### RESIDUAL STRESSES

I have completely omitted any discussion of the role played by residual stresses in contact fatigue. Evidence thus far is inconclusive but suggests a beneficial effect of residual compressive stresses and detrimental effects of tensile residual stresses in the circumferential direction on bearing inner rings (refs. 44 and 64 to 66). Since many of the tests concerning the effects of residual stress on contact fatigue have been conducted at loads sufficient to cause general plastic deformation, it is difficult to appraise the results. The development of residual stresses during contact fatigue testing has been related to the precision elastic limit and hardness difference effects (refs. 33, 34, and 67), and some differences in contact fatigue life have been attributed to the residual stresses observed. Since methods used to obtain variations in residual stress are usually accompanied by significant differences in microstructure, it has not been possible thus far to resolve the separate influence of residual stress in contact fatigue (ref. 68).

#### SUMMARY

The mechanism of contact fatigue is perceived at the macroscopic level as the competition of six modes of crack nucleation and propagation as follows:

- (1) Inclusion origin (at a nonmetallic inclusion or unconfirmed site)
- (2) Subcase fatigue (also known as case crushing, seen only in case-hardened parts)
- (3) Surface origin
  - a. Point surface origin
  - b. General surface damage
  - c. Surface inclusion origin
- (4) Geometric stress concentration



- (5) Peeling (also called superficial pitting or frosting)
- (6) Section fracture

At the scale of the first detectable crack and smaller, the mechanism of crack nucleation and the role of stress-induced microstructural alterations is not well understood. Light etching and dark etching regions formed as a result of cyclic contact stress are useful in disclosing the distribution of unidirectional and reversed shear stresses resulting from contact geometry and revealing the locations of surface or subsurface regions of locally intensified shear stress.

Gross sliding promotes surface-origin modes of fatigue, and it is likely that tangential forces associated with such sliding or microscopic slip within the contact area are the primary cause of surface crack formation. The tendency to surface or subsurface modes of crack initiation is governed by the relative frequency-severity distribution of surface and subsurface discontinuities or stress raisers and is modified by temperature and the chemical activity of the lubricant and/or atmosphere as well as EHD film effects.

Crack propagation is affected by the lubricant and atmosphere from the earliest stages of surface-origin modes of damage and as soon as cracks from subsurface origins reach the surface. Low viscosity at the operating temperature and a chemically aggressive lubricant promote rapid crack propagation in a manner consistent with an hydraulic pressure propagation hypothesis. Alternative mechanisms involve Rebinder or chemical reaction effects.

Crack initiation subsurface is caused by reversed shear stresses and primary subsurface propagation is in shear. Secondary cracking toward the surface is primarily due to the modification of the subsurface stress distribution by the previous cracks. Tensile forces from tangential forces aid in the nucleation and propagation of surface-origin fatigue cracks, but the initial direction of such cracks also indicates that reversed shear stresses at the surface are their primary cause.

Future progress in contact fatigue research will benefit from greater diligence and improved skill in the identification of the various modes of contact fatigue damage encountered in laboratory rig tests or full-scale component tests. A complete accounting of environmental factors is equally important for the correct interpretation of contact fatigue test results and field performance.

#### SUGGESTIONS FOR FUTURE RESEARCH

(1) The role of lubricant composition, additives, contaminants, and degradation products in surface crack initiation should be studied over a range of EHD film/surface roughness ratios and at loads above and below that which will cause bulk plastic deformation.



(2) Crack initiation and propagation under the influence of lubricant base stock or additives might be studied by the methods of Galvin and Naylor (ref. 42). Concentration of additives, temperature, and speed might be the major variables, and noble metal films or imposed electrochemical potential might be used to sort out corrosion, chemical reactions, and Rebinder type effects.

(3) The role of tangential forces without gross sliding should be compared to small degrees of imposed sliding under several EHD film/roughness conditions as they affect peeling, point-surface-origin, and general-surface-damage modes of fatigue.

(4) Studies of surface microcrack growth should be made to determine whether a critical surface crack length for rapid propagation can be related to contact geometry.

(5) Viscosity variations by means of base stock and temperature could be used to evaluate the validity of HPP hypotheses by testing at several rolling velocities after initial surface damage under standardized conditions.

(6) Two fluid tests with careful ultrasonic solvent rinses between exchanges that are made at various time intervals from the beginning of testing should disclose whether lubricant composition exerts major influence upon crack initiation or propagation.

(7) Synthetic surface flaws of measured dimensions might be used to study true contact stress distribution under mixed EHD-asperity contact conditions.

(8) Since surface texture is nearly always altered by mechanical or chemical effects during testing or service, measurements of surface texture should be made at intervals during tests in which it is an important variable.

(9) A budgeted effort to identify modes of fatigue damage should be planned in every contact fatigue study. Relevant details about material composition, microstructure, surface finishing, cleaning, and other details of pretest condition should be provided. Bulk oil inlet and specimen temperatures should be reported. Photographic documentation of typical fatigue damage and of surface conditions (pretest and post-test) should be provided. Where possible two levels of contact stress should be used for testing or two modes of fatigue damage studied.

(10) Tests should be conducted to develop design criteria to avoid subcase fatigue under combined contact fatigue and flexure, press fit, or residual stresses.

(11) Optimization of core microstructure, hardness, core carbon content, and case depth carbon gradient should be attempted for case-hardened bearings, gears, cams, etc. where the economics of machining and other processing costs are less critical than the achievement of maximum performance.

(12) High-frequency ultrasonic inspection using focused beam techniques should be used to study the origin and propagation of subsurface inclusion origin and subcase fatigue damage.

In view of the competition among the various modes of contact fatigue damage, work is needed on the relationship of load (or stress) and life for each fatigue mode. We lack sufficient data at low levels of contact stress to extrapolate results run at very high stress levels with confidence. Chemically aggressive lubricants of low viscosity can be used at appropriate film thickness-surface roughness ratios to keep test times within practical limits. The application of a mathematical model of contact fatigue (ref. 69) will depend upon such information in more complete form than is presently available.

#### ACKNOWLEDGMENT

I wish to express my thanks to The Timken Roller Bearing Company for their permission to present this paper and to include information from experimental laboratory programs and field experience with their bearing and steel products. Also included are results from the work of engineers and technicians in the Physical Laboratories, the Metallurgical Department of the Steel and Tube Division, and at Timken Research. Specifically I acknowledge with thanks the help of R. Widner, H. Burrier, K. Barnes, J. Wolfe, G. Wills, and D. Keener, whose metallographic work and ideas were essential to the preparation of this paper.

#### DISCUSSIONS

**A. Seireg (University of Wisconsin, Madison, Wisconsin)**

Dr. Littmann is to be commended for a valuable and able presentation on the factors influencing fatigue of rolling contacts. Absent from this paper, however, is any mention of the influence of microgeometry and traction on surface failure. Specific reference is made to the recent paper by P. H. Dawson and the discussion by K. L. Johnson and R. J. Pometroy on the paper (ref. 70). It would be of interest to get Dr. Littmann's reactions to the views presented in this paper.

**E. V. Zaretsky (NASA Lewis Research Center, Cleveland, Ohio)**

In recent years Dr. Littmann has been in the forefront of attempting to categorize the rolling-element fatigue phenomena and to bring some order out of apparent chaos. For his efforts, Dr. Littmann should be commended by the technical community. However, what is presented in his paper must be taken as but a first step toward the final goal of defining in detail the rolling-element fatigue phenomena. It is in this light that the following comments are offered.

In references 71 and 72, the maximum orthogonal shearing stress was assumed to be the critical stress instrumental in the rolling-element fatigue process. Dr. Littmann in his paper also discusses and implies the orthogonal shearing stress as the critical stress in causing fatigue. This assumption is reasonably valid if one considers that the maximum reversal of shearing stress occurs on the orthogonal plane. However, metallurgical investigation (ref. 30) of failed rolling-element specimens suggests that the maximum shearing stress, which is on a  $45^\circ$  plane, is the most critical stress. Peak amplitude of the maximum shearing stress is greater than that of the maximum orthogonal shearing stress. Variation of shearing stress on a  $45^\circ$  plane and the orthogonal shearing stress are illustrated and compared in figure 27.

A great deal of significance has been placed on what stresses are the critical ones in rolling-element fatigue. Dr. Littmann outlines the work performed on microstructural alterations in the fatigue process. Most of these data show that the altered regions lie on planes oriented at some angle to the surface beginning at some depth below the rolling-element surface. All of these studies and the conclusions derived therefrom are based on calculated stresses for nonlubricated, two-dimensional Hertzian loading which is not completely valid for the experimental situation, i.e., lubricated full-scale bearings. Until more exact stress calculations for lubricated full-scale bearings are available, it appears meaningless to speculate as to what stresses are causing the observed structural change. It can be concluded, however, that a structural change occurs in the subsurface zone of maximum resolved shearing stresses without any regard to which shearing stress is involved.

It has been suggested in the discussion of reference 33 that there is a threshold energy input below which the structural transformation is not likely to occur. At higher temperatures, more thermal energy is being

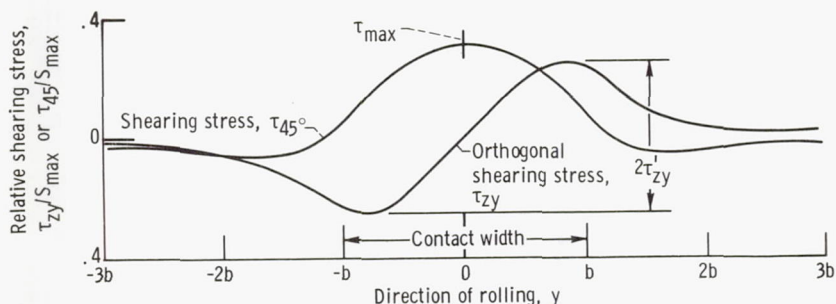


FIGURE 27.—Variation of shearing and orthogonal shearing stresses on plane at depth below rolling surface.



put into the specimen, so that it requires much less input from the cyclic stressing to accomplish the transformation or alteration. Martin et al. (ref. 28) mention that considerable tempering occurs during the hundreds of hours required for testing. Since the microstructural change apparently results from a thermo-mechanical process, there is probably some tempering occurring in the microstructure. As implied by Dr. Littmann, there appears to be no relation between the altered structure and rolling-element fatigue.

Residual compressive stresses are induced in a rolling element during operation. The magnitude of these compressive residual stresses appears to be related to the difference in hardness between the rolling elements such as the balls or rollers and the raceway for through-hardened materials such as 52100. It was found in tests (refs. 73 and 74) that the maximum induced compressive residual stress occurred at a depth which roughly corresponded to the depth of the maximum orthogonal shearing stress. This correlation suggests that the orthogonal shearing stress is that which is instrumental in inducing the residual stress.

The question remains whether the orthogonal stress is, in turn, modified by the residual stress buildup. If a two-dimensional model is considered where it is assumed that the residual stress occurs only in the direction of rolling, a simple analysis indicates that this residual stress is added to any principal Hertzian stress distribution caused by two bodies in contact. A further analysis indicates a significant change in the maximum shearing stress but no change in the maximum orthogonal shearing stress.

Referring to table 7, relative lives were calculated based on residual stresses  $S_r$  at the depth of maximum shearing stress for the bearings of

TABLE 7.—*Relative Predicted Fatigue Lives<sup>a</sup>*

Component hardness difference <sup>b</sup> range, $\Delta H$ , Rockwell C	Average compressive residual stress value, ksi	Relative predicted 10-percent life from residual stress <sup>c</sup>	Relative experimental 10-percent life <sup>c</sup>
-2.0 to -1.0	6	0.06	0.07
-1.0 to 0	36	.19	.23
0 to 1.0	--	----	----
1.0 to 2.0	69	1.00	1.00
2.0 to 3.0	38	.23	.83
3.0 to 4.0	6	.06	.60

<sup>a</sup> Based on residual stress measurements of 207-size deep-groove ball bearing inner races for various values of component hardness difference  $\Delta H$

<sup>b</sup> Ball hardness minus race hardness, Rockwell C scale

<sup>c</sup> Relative to value at  $\Delta H$  range 1.0 to 2.0

references 73 and 74 considering that the calculated maximum shearing stress  $(\tau_{\max})_r$  was modified in accordance with the following equation:

$$(\tau_{\max})_r = \tau_{\max} - \frac{1}{2}(\pm S_r)$$

The positive or negative sign before  $S_r$  indicates a tensile or compressive residual stress, respectively. These results are compared with the relative experimental lives. As can be seen, there is a qualitative correlation between the relative predicted lives and the relative lives determined experimentally. Since the residual stresses affect the value of the shearing stress on a  $45^\circ$  plane and not the orthogonal shearing stresses, and, because there is a correlation between induced residual stress and fatigue life, it must be concluded that the prime stress responsible for affecting fatigue life is the maximum shearing stress. It should be pointed out however that, where there are material defects or inclusions, any one of the maximum shearing stresses can be responsible in a competitive way for inducing an incipient crack or failure.

Dr. Littmann references Greenert (ref. 9) to support the premise that the orthogonal shearing stress is that stress which is critical in rolling-element fatigue. The work of Greenert not only did not consider the differences in stressed volume due to the different geometries used for this correlation, but also did not consider plastic deformation that occurred at the stress levels at which he performed his tests. This plastic deformation could significantly change the stresses on which the conclusions were based.

The discussor must disagree with Dr. Littmann on the theory of hydraulic crack propagation. Too often when a speculation or theory is repeated frequently enough, it takes on the aura of fact. This, in the discussor's opinion, is what has happened to the hydraulic crack propagation theory. In 1936, S. Way (ref. 40) suggested that the cause of fatigue was due to the lubricating fluid penetrating cracks in the surfaces of bodies in contact. Under compressive loading, the hydraulic pressures developed within the crack were assumed sufficient to propagate the crack into a network that resulted in a pit or spall. Way further concluded that, with no oil, no pitting would occur. H. Styri, in a discussion of reference 40 (pp. A110-A114), indicated that locomotive wheels that run without lubrication on rails can pit or spall in a similar manner to a rolling-element bearing. Additionally Styri extended a premise held by others, that crack propagations begin below the rolling-element surface. Contemporary research, of course, has indicated that fatigue can occur at the surface. However, Dr. Littmann performs an injustice to the concept of surface crack initiation and propagation by associating it with the hydraulic theory of propagation. It is the discussor's opinion that the shearing stresses present at the tip of a crack would probably be more than sufficient to propagate the crack into a



network to form a fatigue spall without looking to the presence of a fluid or the lack thereof to explain this phenomenon.

Tests reported in reference 75 have indicated that specimens tested at atmospheric pressure and at the approximate lubricant vapor pressure with two different lubrication methods using a super-refined naphthenic mineral oil as a lubricant produced insignificant differences in fatigue life, deformation, and wear. These results suggest that the oxygen content of the atmosphere plays a very insignificant role in crack initiation in rolling contact. With regard to crack propagation, the discussor believes that this is a moot question.

Types of high-temperature failure modes were discussed in references 76 and 77. These failure modes can be generalized to any operating temperature based on film parameter

$$\Lambda = \frac{h}{\sigma}$$

where the composite surface roughness,  $\sigma$ , is

$$\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2}$$

and  $h$  is the minimum film thickness;  $\sigma_1$  and  $\sigma_2$  are the root mean square roughnesses of the two surfaces in contact. At values of  $\Lambda$  less than 1, surface smearing or deformation accompanied by wear will occur. When  $\Lambda$  is between 1 and 1.5, surface distress will occur. In this region, surface glazing and spalling will occur. References 76 and 77 attribute the glazing phenomenon to marginal EHD film thickness. Under marginal EHD lubrication, high tangential forces can be induced which will tend to relocate the maximum shearing stresses closer to the surface (ref. 4).

For values between 1.5 and 3, some glazing of the surface occurs. However, this glazing will not impair bearing operation or result in a pitting phenomenon. At values of 3 or greater, minimal wear can be expected with no glazing. Where  $\Lambda$  is greater than 4, full EHD separation of the surfaces can be expected.

It is interesting to note that tests with 25-mm bore bearings and a synthetic paraffinic oil without an anti-wear additive at 600° F produced some surface glazing. For the same tests with the lubricant having an anti-wear additive, no glazing was apparent. It is conceivable that anti-wear additives will reduce the tangential forces under marginal EHD conditions and thus increase bearing life.

**A. J. Gentile (AC Electronics, Division of General Motors Corporation, Milwaukee, Wisconsin)**

For several years I have been engaged in work devoted to lubrication and surface effects in contact loaded applications. There is, I feel, one area which is completely neglected in these investigations and that is



the metallurgical effect of the surface preparation. Although extreme care is taken to reproduce specimens, there is no effort made to investigate the surface condition on either a sample or 100-percent basis. As a result, the data can be off by a significant amount.

In a paper being co-authored by D. McCormick of the New Departure Hyatt Division of GMC and myself, we will present bearing endurance results illustrating this point. In these tests, we varied the surface preparation on a group of bearings to include various mechanical and chemical finishing techniques and found a 2:1 difference in the  $B_{10}$  life between the lots with the same lubricant. The number of bearings involved was sufficient to lend statistical significance to these results. Metallurgical examinations reveal considerable difference in surface properties depending upon finishing treatments utilized. These results emphasize the point that any testing involving one metal in contact with another, with or without lubrication, should definitely include a metallurgical assessment of the surface.

**S. Allen (MIT Instrumentation Laboratory, Cambridge, Massachusetts)**

In examining data obtained from bearings whose surfaces were finished by different techniques, such as those given in table 6 of Dr. Littmann's paper, it is important to realize that different machining and finishing techniques often result in different near-surface properties which in turn may influence test results. Description of surface texture or topography, therefore, is not sufficient. Other aspects such as residual stress and details of near-surface microstructure also need defining. Residual stress information is occasionally given, but information on the near-surface structure, which may be independent of it, is scant. Data relating to manufacturing effects on near-surface structure were published by Scott et al. (ref. 44). These, by section electron microscopy, indicated metallographic modification to a depth of about 80  $\mu$ in. from grinding and a shallower one due to polishing. The modified microstructure was found detrimental by fatigue tests.

Working with experimental gyro spin-axis ball bearings, a near-surface metallographic modification at least 10  $\mu$ in. deep and due to the post-grind finishing operation was demonstrated by one of us (ref. 78), using a surface etching technique. Subsequent work by the other, using a sectioning technique, has confirmed the conclusions on other races, finished by essentially similar methods. The surface modification of gyro bearings is displayed by etching in figure 28 and by sectioning in figure 29. Similar micrographs for races finished by lapping techniques, which do not result in discernable near-surface structure changes, are shown in figures 30 and 31. The race materials of figures 28 and 30 are 440C and those of the other two 52100. In the report cited (ref. 78), the modification was definitely traced to the finishing steps of honing used on the

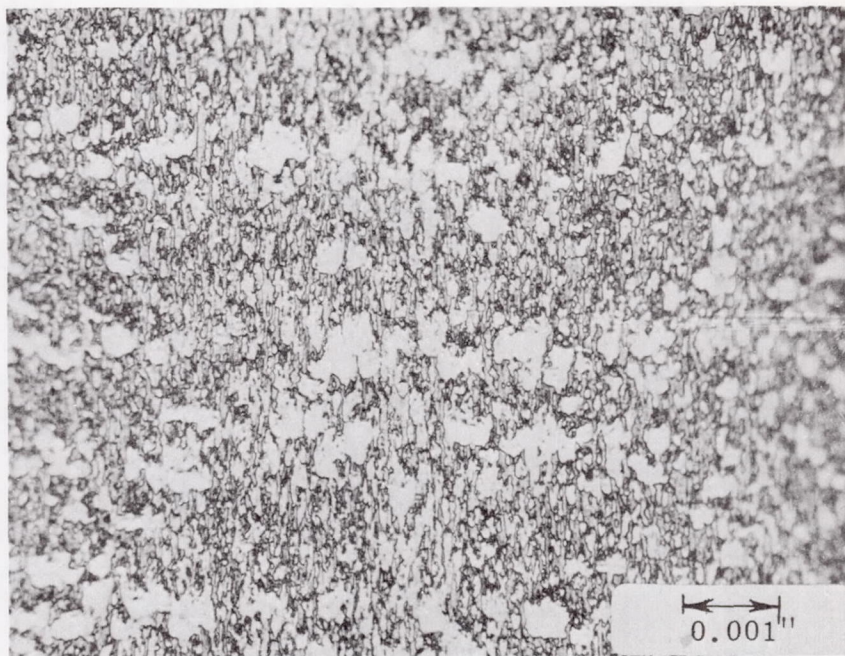


FIGURE 28.—Microstructure of 440C gyro bearings distorted by the finishing operation. Etched race groove, 500 $\times$ .

race grooves. Subsequent work again correlated the most severe changes with a similar honing technique in some other lots of gyro bearings. Gyro bearings are so lightly loaded (180 ksi max. Hertz) that they are not fatigue limited, thus Scott et al. (ref. 44) data do not apply. The highly distorted structure is believed to be undesirable in these bearings (ref. 78), but statistically valid data to confirm this under running conditions do not exist.

The high degree of deformation observed in these hardened and tempered steels ( $\geq R_c60$ ) is probably possible only with triaxial stress deformation conditions and resembles that seen after rubbing contact experiments (ref. 79). It is interesting to speculate whether such microstructural deformation is perhaps connected with some of the peeling damage observed by Dr. Littmann in fatigue-limited bearings. The depth of that effect is shallow, as is the modified microstructure observed (ref. 78), and the shallow crack angles cited by Dr. Littmann may correlate with the microstructural deformation (see fig. 29). Finally, the most severe microstructural deformations observed (ref. 78) were from some of the techniques that achieved the smoothest finishes, suggestive of Littmann's observation that smoother surfaces suffer more severe peeling damages. A wider usage of near-surface metallographic information is indicated.



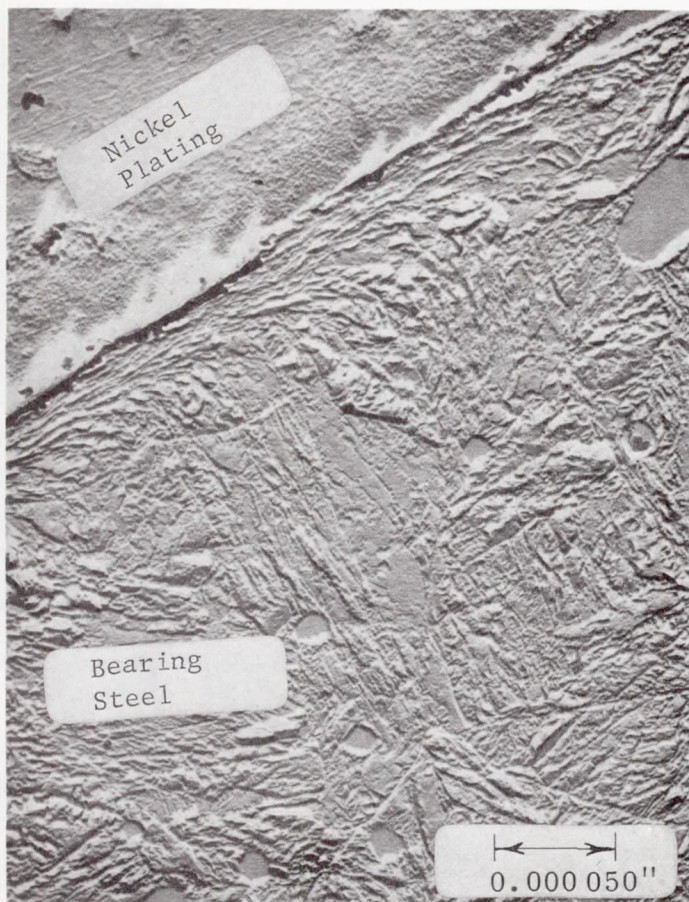


FIGURE 29.—52100 gyro bearing finished by a technique similar to that of figure 28. Transverse section showing nickel plated race groove, 12 500 $\times$ . Light streak on nickel at steel interface is an artifact.

#### LECTURER'S CLOSURE

The author wishes to express his thanks to the discussors for their comments and the valuable information in their contributions. Replying first to Dr. Seireg, perhaps I did not review in sufficient detail the role of tractive forces and microgeometry. Both of these subjects were discussed in the treatment of surface-initiated contact fatigue cracks (surface-origin damage and general surface distress); and of the six major variables of importance for control of surface-origin modes of contact fatigue damage, surface microtopography was the first one discussed. The fatigue life test results in table 6 were not treated in detail because



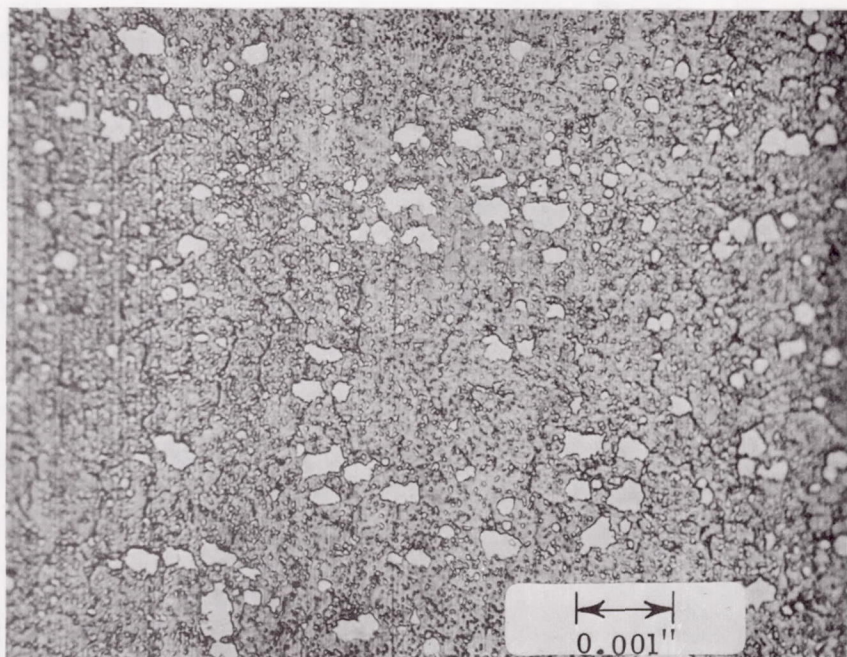


FIGURE 30.—440C gyro bearing finished by technique that does not distort microstructure. Etched race groove, 500X.

it would have added excessively to the length of an already long manuscript. A close study of table 6 reveals some important facts worthy of additional study. At the same ratio of film thickness to composite roughness (Test 3 vs Test 5), a significantly longer life was obtained with one surface honed and more roller fatigue damage occurred when both surfaces were ground. The cups (outer races) were honed and identical in all tests reported in table 6, and no cup failures occurred. In Test 4 vs Test 6, the  $h_a/\sigma_1$  ratio does not explain the significantly poorer life for the combination of lower composite roughness. Clearly more work is needed on surface microgeometry effects.

The observations described in Dawson's recent paper (ref. 70) are consistent with most of the results seen in our laboratories for carburized and hardened steels. I believe that most of the evidence indicates that shallow cracks running parallel to the surface usually intersect the surface at some location as seen in figure 20. The tendency to shallow propagation parallel to the surface is one of the characteristics of peeling damage. The early appearance of surface cracks reported by Dawson confirms the results of Reichenbach et al. (ref. 46) and seems to characterize the conditions that develop when EHD film thickness is small compared to surface roughness.

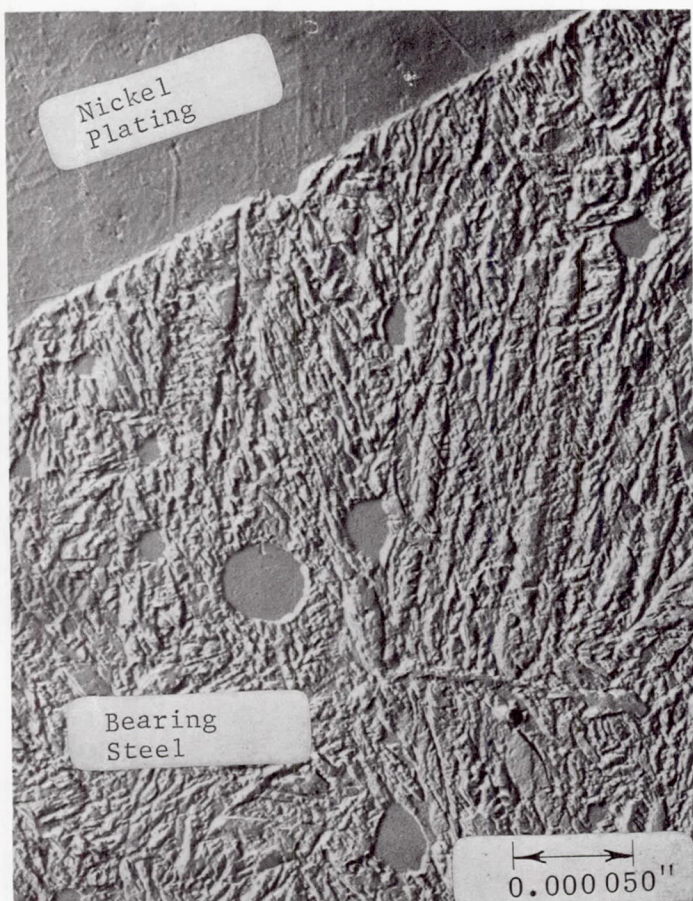


FIGURE 31.—52100 gyro bearing finished by another technique which also does not distort microstructure. Transverse section, 12 500 $\times$ .

The quantitative role of surface traction is not fully understood, but the evidence is clear that sliding superimposed on rolling contact promotes surface-origin damage. The influence of varied tractive force without gross sliding is not known. It seems clear that the local surface stresses at asperities or discontinuities in the composite contact surface are greater than the nominal computed stresses, even when tangential forces are taken into account, and these are the stresses that cause crack initiation.

The author thanks Mr. Zaretsky for his added observations and com-



ments. It is true that his analysis of the observed influence of residual stresses induced by hardness differences between contact surfaces suggests that the maximum shear stress range may be critical under some circumstances. It is true that plastic deformation in reference 9 modified the computed stresses. Nevertheless, the cracks seen in Greenert's test specimens correlated in direction and depth with the computed orthogonal shear stresses. It would appear from these data and those from microstructural alterations that there are differing conditions under which one or the other shear stress is dominant in its effect. The major conclusion with respect to critical stresses, in the author's opinion, is that stress concentrations at internal or surface discontinuities produce the locally highest stresses that cause fatigue crack initiation.

Opinions on the merit of the hydraulic pressure hypothesis can be supported by a variety of observations that are consistent with such a mechanism. In the review of the author's observations and published information on this subject, it was intended simply to present the relevant information for consideration. Some critical experiments are needed before this hypothesis can be laid aside or given credence.

Mr. Zaretsky's information on the experiments in hydrocarbon vapor does not rule out a possible role of oxygen when water is present, as it often is in many machine environments. The author agrees that, except for films that affect boundary lubrication, the major chemical effects in contact fatigue are probably seen in the propagation rather than in the origin of fatigue cracks.

Under low values of  $A$ , one would expect anti-wear additives to influence surface damage, as Mr. Zaretsky reported. Indeed, all aspects of lubricant chemistry seem to be more critical when the EHD film thickness is less than the composite surface roughness. Yet, as shown by the tests described in tables 4 and 5 (illustrated in figures 12 and 13), the chemical properties of the lubricant affect propagation from subsurface-inclusion origins in a manner which has practical significance, too.

Mr. Gentile and Dr. Allen call attention to a much neglected aspect of contact fatigue research. The metallurgical investigation and documentation of surface microstructure is of particular importance when operating conditions promote surface-origin modes of contact fatigue. Differences in techniques used for surface preparation are reflected in both the microtopography and the microstructure. A correct interpretation requires equal attention to the documentation of both characteristics. The fine micrographs that Dr. Allen has contributed illustrate the dramatic differences that can be seen in surfaces which might appear identical by microtopographic measurements.

In the studies made in The Timken Company laboratories, we customarily employ macroscopic etching techniques and electroplated taper sections to examine the microstructure of the contact surfaces in test



bearings. The typical condition of ground or honed surfaces of the tapered roller bearings used in the studies cited in work by the author and his associates are similar to the lapped conditions illustrated in Dr. Allen's figures 30 and 31.

The early stages of peeling damage shown in figure 25 seem to be associated with a layer of altered microstructure about 100  $\mu$ n. deep. This layer developed from the repeated stresses of test and was not present prior to the test. We have not associated the initiation of peeling with a surficial layer of altered microstructure due to the method of surface finishing. The author's comments on the tendency to greater peeling damage on the smoother surface when contacting surfaces have different roughness is simply a report of what has been observed. Certainly a systematic study of the tendency to peeling damage should be made on smooth surfaces having the types of surface microstructures illustrated by Dr. Allen.

#### REFERENCES

1. JOHNSON, L. G.: Statistical Treatment of Fatigue Experiments. Elsevier Publ., Amsterdam, 1964.
2. WEIBULL, W.: Fatigue Testing and Analysis of Results. Pergamon Press, Oxford, 1961.
3. LITTMANN, W. E.; AND MOYER, C. A.: Competitive Modes of Failure in Rolling Contact Fatigue. SAE Paper 620A, January 1963.
4. SMITH, J. O.; AND LIU, C. K.: Stresses Due to Tangential and Normal Loads on an Elastic Solid with Application to Some Contact Stress Problems. J. Appl. Mech., vol. 20, no. 2, 1953, pp. 157-166.
5. SACHS, G.; SELL, R.; AND BROWN, W. F., JR.: Tension Compression and Fatigue Properties of Several Steels for Aircraft Bearing Applications. Proc. ASTM, vol. 59, 1959, p. 635.
6. ZARETSKY, E. V.; AND ANDERSON, W. J.: Rolling Contact Fatigue Studies with Four Tool Steels and a Crystallized Glass Ceramic. Trans. ASME, Jour. Basic Engrg., vol. 83, no. 4, 1961, pp. 603-612.
7. RADZIMOVSKY, E. I.: Stress Distribution and Strength Condition of Two Rolling Cylinders Pressed Together. Univ. of Illinois Eng. Exp. Sta., Bull. 408, 1953.
8. MOYAR, G. J.; AND MORROW, J.: Surface Failure of Bearings and Other Rolling Elements. Univ. of Illinois Eng. Exp. Sta., Bull. 468, 1964.
9. GREENERT, W. J.: The Toroid Contact Roller Test as Applied to the Study of Bearing Materials. Trans. ASME, J. Basic Engrg., vol. 84, 1962, p. 181.
10. JOHNSON, K. L.; AND JEFFRIES, J. A.: Plastic Flow and Residual Stresses in Rolling and Sliding Contact. Fatigue in Rolling Contact, Inst. Mech. Engrs., (London) 1963, pp. 54-65.
11. THOMAS, H. R.; AND HOERSCH, V. A.: Stresses Due to the Pressure of One Elastic Solid Upon Another. Univ. of Illinois Eng. Exp. Sta., Bull. 212, 1930.
12. CHINNERY, M.; MCCONNELL, R.; LEOMBRUNO, W.; AND O'BRIEN, J.: Observation of Fatigue Deformation in Contact Loading. ASME Paper 67-DE-53, 1967.
13. TALLIAN, T. E.: Discussion on Topography of Solid Surfaces by J. B. P. Williamson. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, pp. 118-127.
14. ANON.: Friction, Wear, and Lubrication Terms and Definitions. First Revision, Organization for Economic Co-operation and Development, Paris, 1968.

15. MITSUDA, T.: An Investigation of Pitting and Shelling Failures in Rolling Contact. T & A.M. Report No. 286, Univ. of Illinois, 1965.
16. TALLIAN, T. E.: On Competing Failure Modes in Rolling Contact. ASLE Trans., vol. 10, 1967, pp. 418-439.
17. MARTIN, J. A.; AND EBERHARDT, A. D.: Identification of Potential Fatigue Nuclei in Rolling Contact Fatigue. Trans. ASME, J. Basic Engrg., 1967, pp. 932-942.
18. DENNING, R. E.; AND RICE, S. L.: Surface Fatigue Research with the Geared Roller Test Machine. SAE Paper 620B, 1963.
19. WREN, F. J.; AND MOYER, C. A.: Modes of Fatigue Failures in Rolling Element Bearings. Proc. Inst. Mech. Engrs., vol. 179, pt. 3D, 1964-65, pp. 236-247.
20. FESSLER, H.; AND OLLERTON, E.: Contact Stresses in Toroids under Radial Loads. British J. Appl. Phys., vol. 8, no. 10, 1957, p. 387.
21. BARWELL, F. T.; AND SCOTT, D.: The Investigation of Unusual Bearing Failures. Proc. Inst. Mech. Engrs., vol. 180, pt. 3K, 1965-66, pp. 277-297.
22. SCOTT, D.; LOY, B; AND MILLS, G. H.: Metallurgical Aspects of Rolling Contact Fatigue. Proc. Inst. Mech. Engrs., vol. 181, pt. 30, 1966-67, pp. 94-103.
23. LITTMANN, W. E.; AND WIDNER, R. L.: Propagation of Contact Fatigue from Surface and Subsurface Origins. Trans. ASME, J. Basic Engrg., vol. 188, no. 3, 1966, pp. 624-636.
24. WIDNER, R. L.; AND WOLFE, J. O.: Analysis of Tapered Roller Bearing Damage. ASM Tech. Report No. C7-11.1, 1967.
25. DAVIES, W. J.; AND DAY, K. L.: Surface Fatigue in Ball Bearings, Roller Bearings and Gears in Aircraft Engines. Fatigue in Rolling Contact, Inst. Mech. Engrs. (London), 1963, pp. 23-40.
26. MURRAY, J. D.; AND JOHNSON, R. F.: The Effect of Inclusions on the Properties of 1% C-Cr Bearing Steels. Special Report 77, The Iron and Steel Institute, 1963, pp. 104-109.
27. LYNE, C. M.; AND KASAK, A.: Effect of Sulfur on the Fatigue Behavior of Bearing Steel. Trans. ASM, vol. 61, no. 1, 1968, pp. 10-13.
28. MARTIN, J. A.; BORGESSE, S. F.; AND EBERHARDT, A. D.: Microstructural Alterations of Rolling Bearing Steel Undergoing Cyclic Stressing. Trans. ASME, J. Basic Engrg., vol. 88, no. 3, 1966, pp. 555-567.
29. SCHLICHT, H.: Strukturelle Änderungen in Wälzelementen. Wear, vol. 12, 1968, pp. 149-163.
30. CARTER, T. L.: A Study of Some Factors Affecting Rolling Contact Fatigue Life. NASA TR-R-60, 1960.
31. STYRI, H.: Fatigue Strength of Ball Bearing Races and Heat Treated 52100 Steel Specimens. Proc. ASTM, vol. 51, 1951, p. 682.
32. O'BRIEN, J. L.; AND KING, A. H.: Electron Microscopy of Stress-Induced Structural Alterations Near Inclusions in Bearing Steels. Trans. ASME, J. Basic Engrg., vol. 88, no. 3, 1966, pp. 568-572.
33. BUSH, J. T.; GRUBE, W. L.; AND ROBINSON, G. H.: Microstructural and Residual Stress Changes in Hardened Steel Due to Rolling Contact. Trans. ASM, vol. 54, 1961, pp. 390-412.
34. GENTILE, A. J.; JORDAN, E. F.; AND MARTIN, A. D.: Phase Transformations in High Carbon-High Hardness Steel under Contact Loads. Trans. Met. Soc. AIME, vol. 233, 1965, pp. 1085-1093.
35. LITTMANN, W. E.; WIDNER, R. L.; STOVER, J.; AND WOLFE, J. O.: The Role of Lubrication in the Propagation of Contact Fatigue. Trans. ASME, J. Lub. Tech., 1968, pp. 89-100.
36. ANON: Steel. March 5, 1962, p. 74.



37. SCHATZBERG, P.; AND FELSEN, I. M.: Influence of Water on Fatigue Failure Location and Surface Alteration During Rolling Contact Lubrication. ASME Paper 68-Lub-11.
38. FOORD, C. A.; HINGLEY, C. C.; AND CAMERON, A.: Pitting of Steel Under Varying Speeds and Combined Stresses. ASLE Preprint, presented at ASME-ASLE Lubrication Conf. (Atlantic City, New Jersey), October 1968.
39. WAY, S.: Pitting Due to Rolling Contact. Trans. ASME, vol. 57, 1935, pp. A49-A58.
40. WAY, S.: Gear Tooth Pitting. The Electric J., April 1936.
41. DUQUETTE, D. J.; AND UHLIG, H. H.: Effect of Dissolved Oxygen and NaCl on Corrosion Fatigue of 0.18% C Steel. Trans. ASM, vol. 61, no. 3, 1968, pp. 449-456.
42. GALVIN, G. D.; AND NAYLOR, H.: Effect of Lubricants on the Fatigue of Steel and Other Metals. Proc. Inst. Mech. Engrs., vol. 179, pt. 3J, 1964-65, pp. 56-70.
43. SPITZIG, W. A.; AND WEI, R. P.: A Fractographic Investigation of the Effect of Environment on Fatigue Crack Propagation in an Ultra High Strength Steel. Trans. ASM, vol. 60, no. 3, 1967, pp. 279-288.
44. SCOTT, R. L.; KEPPEL, R. K.; AND MILLER, M. H.: The Effect of Processing-Induced Near Surface Residual Stress on Ball Bearing Fatigue. Rolling Contact Phenomena, Elsevier Publ. (Amsterdam), 1962, pp. 301-315.
45. BORGESE, S.: An Electron Fractographic Study of Spalls Formed in Rolling Contact. Trans. ASME, J. Basic Engrg., vol. 89, 1967, pp. 943-948.
46. REICHENBACH, G. S.; SYNIUTA, W. D.; AND WALTER, D.: An Electron Microscope Study of Rolling Contact Fatigue. ASLE Trans., vol. 8, no. 3, 1965, p. 217.
47. PEDERSON, R.; AND RICE, S. L.: Case Crushing in Carburized and Hardened Gears. SAE Trans., vol. 69, 1961, pp. 370-380.
48. NIEMANN, G.; RETTIG, H.; AND BÖTSCH, H.: The Effect of Different Lubricants on Pitting Resistance of Gears. Proc. Inst. Mech. Engrs., vol. 179, pt. 3D, 1964-65, pp. 192-200.
49. DAWSON, P. H.: Further Experiments on the Effect of Metallic Contact on the Pitting of Lubricated Rolling Surfaces. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965, pp. 95-100.
50. WILLIAMSON, J. B. P.: Topography of Solid Surfaces. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, pp. 85-142.
51. KELLEY, B. W.: The Importance of Surface Temperature to Surface Damage. Handbook of Mech. Wear, D. Lipson and L. V. Colwell, eds., Univ. of Michigan Press, ch. 8, 1961, p. 155.
52. KOVED, I.: The Effect of Three Mineral Base Oils on Roller Bearing Fatigue Life. ASLE Trans., vol. 9, 1966, pp. 222-228.
53. STERNLICHT, B.; LEWIS, P.; AND FLYNN, P.: Theory of Lubrication and Failure of Rolling Contacts. Trans. ASME, J. Basic Engrg., vol. 83, 1961, p. 213.
54. ANDERSON, E. L.; AND KU, P. M.: Rolling Contact Fatigue and Wear Characteristics of Three Oils as Determined by a Three Ball/Cone Tester. Proc. USAF-SWRI Turbine Engine Lubrication Conf., 1967, pp. 98-108.
55. KOVED, I.: A Comparison of Fatigue Test Techniques for Gas Turbine Oils. SAE Paper 680322, 1968.
56. LUCAS, G.: Surface Damage to Forged Steel Rolls and Its Prevention. Proc. Inst. Mech. Engrs., vol. 179, pt. 3D, pp. 271-280.
57. SIMPSON, F. F.: Failure of Rolling Contact Bearings. Proc. Inst. Mech. Engrs., vol. 179, pt. 3D, 1964-65, pp. 248-270.
58. MCKELVEY, R. E.; AND MOYER, C. A.: The Relation Between Critical Maximum Compressive Stress and Fatigue Life Under Rolling Contact. Fatigue in Rolling Contact, Inst. Mech. Engrs., 1963.



59. MOYER, C. A.; AND NEIFERT, H. R.: A First Order Solution for the Stress Concentration Present at the End of Roller Contact. ASLE Trans., vol. 6, 1963, 324-336.
60. SHAW, M. C.: Friction and Wear Interdisciplinary Workshop. NASA TMX-52748, 1970.
61. GOHAR, R.; AND CAMERON, A.: The Mapping of Elastohydrodynamic Contacts. ASLE Trans., vol. 10, 1967, p. 215.
62. MOORE, C. C.; PERKINS, P. A.; AND SMEATON, D. A.: Flange Mounted Ball Bearing Fatigue Tests. Trans. ASME, J. Basic Engrg., vol. 89, 1967, pp. 919-931.
63. JONES, A. B.: Metallographic Observations of Ball Bearing Fatigue Phenomena. Proc. ASTM, vol. 46, p. 1.
64. GENTILE, A. J.; AND MARTIN, A. D.: The Effects of Prior Metallurgically Induced Compressive Residual Stress on the Metallurgical and Endurance Properties of Overload Tested Ball Bearings. ASME Preprint 65 WA/CF-7, 1965.
65. HUSTEAD, T. E.: Considerations of Cylindrical Roller Bearing Load Rating Formula. SAE Preprint 569A, September 1962.
66. ALMEN, J. O.: Effects of Residual Stress on Rolling Bodies. Contact Fatigue Phenomena, Elsevier Publ. (Amsterdam), 1962, pp. 400-424.
67. ZARETSKY, E. V.; PARKER, R. J.; AND ANDERSON, W.: Component Hardness Differences and Their Effects on Bearing Fatigue. Trans. ASME, J. Lub. Tech., vol. 89F, 1967, pp. 47-62.
68. ERICKSON, J. A.: Improvement of Bearing Steels Through Heat Treatment and Steel Mill Processing. ASM Tech. Report N.C. 6-5.1, 1966.
69. CHIU, Y. P.; TALLIAN, T. E.; MCCOOL, J. I.; AND MARTIN, J. A.: A Mathematical Model of Spalling Fatigue Failure in Rolling Contact. ASLE Trans., vol. 12, 1969, p. 106.
70. DAWSON, P. H.: Rolling Contact Fatigue Crack Initiation in a 0.3 Percent Carbon Steel. Proc. Inst. Mech. Engrs., vol. 183, pt. 1, no. 4, 1968-69.
71. LUNDBERG, G.; AND PALMGREN, A.: Dynamic Capacity of Rolling Bearings. Acta Polytech, Mech. Eng. Ser., vol. 1, no. 3, 1947.
72. LUNDBERG, G.; AND PALMGREN, A.: Dynamic Capacity of Rolling Bearings. J. Appl. Mech., vol. 16, no. 2, 1949, pp. 165-172.
73. REICHARD, D. W.; PARKER, R. J.; AND ZARETSKY, E. V.: Residual Stress and Subsurface Hardness Changes Induced During Rolling Contact. NASA TN D-4456, 1968.
74. ZARETSKY, E. V.; PARKER, R. J.; AND ANDERSON, W. J.: A Study of Residual Stress Induced During Rolling. Trans. ASME, J. Lub. Tech., vol. 91F, 1969, pp. 314-319.
75. REICHARD, D. W.; PARKER, R. J.; AND ZARETSKY, E. V.: Studies of Rolling-Element Lubrication and Fatigue Life in a Reduced Pressure Environment. ASLE Trans., vol. 11, no. 3, 1968, pp. 275-280.
76. SIBLEY, L. B.: Elastohydrodynamic Lubrication. Machine Design, vol. 38, no. 28, 1966, pp. 220-221.
77. GIVEN, P. S.: Lubrication Film Effects on Rolling-Contact Fatigue. Paper presented at Dartmouth College Bearings Conf., Hanover, New Hampshire, September 1966.
78. ALLEN, S.; AND PALMIERI, J. R.: A Metallurgical Modification Caused by Finish Operations on Ball-Bearing Race Grooves. E-2084, MIT Instrumentation Lab., December 1966.
79. WILSON, F.; AND EYRE, T. S.: Metallographic Aspects of Wear. Metals and Materials, vol. 3, 1969, p. 86.

**Page intentionally left blank**

# Effect of Materials—General Background

**E. V. ZARETSKY  
W. J. ANDERSON**

**NASA Lewis Research Center  
Cleveland, Ohio**

A review of state-of-the-art and current advances in rolling-element bearing material technology is presented. Topics include materials, inclusion content, heat treatment, fiber orientation, residual stresses, ausforming, component differential hardness, and material dislocations.

**T**HE ROLLING-CONTACT PHENOMENON, the interaction of two or more bodies in rolling contact, has been the subject of much study, debate, and controversy for several decades. Many unresolved questions currently exist, in addition to numerous new problems which should develop as machine elements become more complex and as loads, temperatures, and speeds increase within these elements. Fatigue is a mode of failure that occurs in rolling elements such as ball and roller bearings. It is a cyclic-dependent phenomenon resulting from repeated stresses under rolling-contact conditions. Fatigue can be affected by many variables, such as rolling speed, load, material, sliding within the contact zone, temperature, contact geometry, type of lubricant, and others. The fatigue failure manifests itself initially as a pit which, in general, is limited in depth to the zone of resolved maximum shearing stresses and in diameter to the width of the contact area.

The phenomenal growth of the aircraft industry subsequent to World War II created unprecedented needs for better design and materials for rolling-element bearings used in subsonic and supersonic aircraft. These needs included bearings for increased temperatures, higher speeds, and greater loads. Coupled with these requirements, and compounding the difficulty, is the higher ratio of thrust to weight which dictated the use of smaller and lighter bearings. Reliability of these bearings became a major consideration due not only to system and mission complexities, but also to the staggering cost aspects involved. Research and development performed by the bearing industry, engine manufacturers, and the Gov-



ernment have provided better bearing materials and manufacturing techniques that have enhanced bearing life and reliability (ref. 1).

## BEARING STEELS

### Through-Hardened Steels

The bearing industry has used SAE 52100 steel as a standard material since 1920. This is a high-carbon chromium steel that also contains small amounts of manganese, silicon, nickel, copper, and molybdenum. This alloy is ductile, strong, and wear resistant. For bearings it is generally air-melted in electric furnaces and has a high degree of cleanliness from rigorous control of the melting process.

A commonly accepted minimum tolerable hardness for bearing components is 58 Rockwell C ( $R_C$ ). At a hardness below this value, brinelling of the bearing races can occur. Since hardness decreases with temperature, conventional bearing materials such as SAE 52100 can be used only to temperatures of about 350° F. Much effort has gone into developing steel alloys suitable for higher temperatures. The addition of elements such as molybdenum, tungsten, chromium, and vanadium promote the retention of hardness at elevated temperatures. Other materials that promote hardness retention are aluminum and silicon.

Molybdenum air-hardening steels were initially formulated for use in tools where cleanliness of structure was of little importance. Continual refinement of metallurgical structure has been necessary to make them suitable for use in bearings. Compositions of some high-temperature alloys and some of the more conventional alloys are given in table 1. Representative hot hardnesses for several of these materials are given in figure 1.

For a corrosive environment, 440C stainless steel should be considered. However, its maximum obtainable hardness is approximately 60  $R_C$ . From the hot-hardness curve of figure 1, the material is too soft for temperatures much above 300° F. In addition, the dynamic capacity is not comparable to conventional SAE 52100.

M-50 and Halmo are sometimes referred to as intermediate high-temperature steels and are usable to about 600° F. Percentages of alloying elements are somewhat lower than the so-called high-temperature steels. Of the intermediate high-temperature steels, M-50 is most widely used. The high-temperature molybdenum-based steels (M-1, M-2, M-10) contain sufficient percentages of chromium, molybdenum, tungsten, and vanadium (in addition to a high carbon content) for use to about 800° F in an oxidizing environment. The hardnesses of M-1 and M-2 are satisfactory to about 900° F, but oxidation resistance becomes marginal at this temperature. An additional alloy that has shown promise for temper-

TABLE 1.—*Through-Hardened Rolling-Element Bearing Steels*

Designation	Max. operating temp. (°F)	Alloying element (% by wt)									
		C	P (max.)	S (max.)	Mn	Si	Cr	V	W	Mo	Co
440C	300	1.03	0.018	0.014	0.48	0.41	17.3	0.14	—	0.50	—
SAE 52100	350	1.00	0.025	0.025	0.35	0.30	1.45	—	—	—	—
MHT*	500	1.03	0.025	0.025	0.35	0.35	1.50	—	—	—	—
Halmo	600	0.65	0.030	0.030	0.27	1.20	4.60	0.55	—	5.2	—
AISI M-50	600	0.80	0.030	0.030	0.30	0.25	4.00	1.00	—	4.25	—
T-1	800	0.70	0.030	0.030	0.30	0.25	4.00	1.00	18.0	—	—
AISI M-10	800	0.85	0.030	0.030	0.25	0.30	4.00	2.00	—	8.00	—
AISI M-1	900	0.80	0.030	0.030	0.30	0.30	4.00	1.00	1.50	8.00	—
AISI M-2	900	0.83	0.030	0.030	0.30	0.30	3.85	1.90	6.15	5.00	—
WB-49	1000	1.07	0.006	0.007	0.30	0.02	4.40	2.00	6.80	3.90	5.2

\* Also contains 1.36% Al

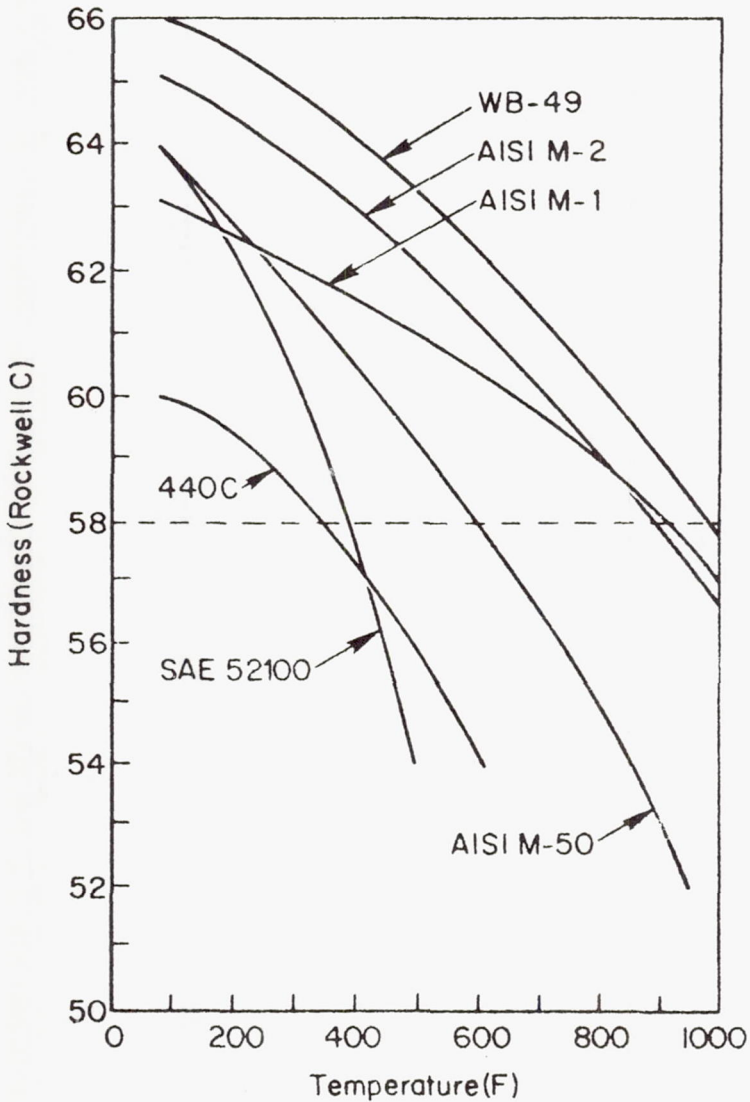


FIGURE 1.—Hot hardness of several through-hardened bearing steels.

atures beyond 900° F is WB-49. Also, tungsten-bearing steel T-1 has been used to some extent in high-temperature bearings (refs. 2 and 3).

Experience has shown that for AISI M-50 and M-1, the dynamic capacity as calculated by AFBMA methods can be exceeded to 600° F. However, experience with the WB-49 material has shown that bearing dynamic capacity should be derated if it is used to 600° F. Regardless



of bearing operating temperature, adequate lubrication must be provided to get satisfactory bearing operation (refs. 4 and 5).

#### **Carburized or Case-Hardened Materials**

Carburized materials are generally used for roller bearings. The materials are characterized by a hardened surface (greater than 0.015-in. thick) and a soft core. The room-temperature hardness of most carburized bearing steels is approximately 58 to 63 R<sub>C</sub>, and they have core hardnesses of 25 to 40 R<sub>C</sub>. As a consequence, these materials are limited to operating temperatures of less than 350° F. A list of carburized materials and their usage is given in table 2.

A hot-hardness curve for commonly-used carburized AISI 4620 is given in figure 2 and is representative of other extensively used carburized steels.

Recent technology has extended the temperature limitations of carburized materials. Three of the high-temperature carburized steels are Bower 315 and Timken CBS 600 and 1000; figure 2 shows two of them. From these curves, CBS 600 appears to be limited to temperatures less than 450° F, and CBS 1000 appears to be limited to approximately 600° F (ref. 6).

#### **Fatigue-Life Comparisons**

There has been a considerable amount of fatigue studies performed comparing the life endurance of various bearing materials (refs. 3 and 7 to 10). However, at the time these data were generated there was incomplete knowledge of the effect of various material and physical variables on rolling-element fatigue such as material heat, hardness, and lubricant type and batch control. As a result, most of these variables were not controlled rendering most of the data useless for the purpose of comparing different materials. The first set of data in the open literature that allowed limited comparison of bearing materials was published by Carter et al. (ref. 11) and is presented in figure 3. Even for these data, which were generated in the NASA spin rig, there is some question as to whether an absolute comparison can be made inasmuch as the absolute hardness difference between the ball specimens and the race cylinder was not always a controlled variable. These data do indicate, however, that the Halmo material may be statistically better than the M-1 and M-50 materials. There is also an indication that the M-50 material is superior to the M-1 material at room temperature. It must be kept in mind, however, that this comparison was made under those conditions specified in figure 3, and a broad generalization may not be valid without additional data. A subjective comparison of the data of reference 3 leads to the conclusion that there may not be any significant life differences at 200° F outer-race temperature between SAE 52100, AISI M-1, AISI M-10, and AISI

TABLE 2.—*Carburized Rolling-Element Bearing Steels*

AISI designation	Usage	Alloying element (% by wt)						
		C	Mn	Si	Cr	Ni	Mo	Others
1015	Med	0.13/0.18	0.30/0.60	—	—	—	—	—
1019	Med	0.15/0.20	0.70/1.00	—	—	—	—	—
1020	Low	0.18/0.23	0.30/0.60	—	—	—	—	—
1024	Low	0.19/0.25	1.35/1.65	—	—	—	—	—
1118	Med	0.14/0.20	1.30/1.60	—	—	—	—	Sulfur 0.08/0.13
3310*	High	0.08/0.13	0.45/0.60	—	1.40/1.75	3.25/3.75	—	—
4023	Low	0.20/0.25	0.70/0.90	—	—	—	0.20/0.30	—
4027	Low	0.25/0.30	0.70/0.90	—	—	—	0.20/0.30	—
4118	Med	0.18/0.23	0.70/0.90	—	0.40/0.60	—	0.08/0.15	—
4320	High	0.17/0.22	0.45/0.65	—	0.40/0.60	1.65/2.00	0.20/0.30	—
4422*	Low	0.20/0.25	0.70/0.90	—	—	—	0.35/0.45	—
4620	High	0.17/0.22	0.45/0.65	—	—	1.65/2.00	0.20/0.30	—
4720	High	0.17/0.22	0.50/0.70	—	0.35/0.55	0.90/1.20	0.15/0.25	—
4820	Low	0.18/0.23	0.50/0.70	—	—	3.25/3.75	0.20/0.30	—
5120	Med	0.17/0.22	0.70/0.90	—	0.70/0.90	—	—	—
8620	High	0.18/0.23	0.70/0.90	—	0.40/0.60	0.40/0.70	0.15/0.25	—
9310*	Low	0.08/0.13	0.45/0.65	—	1.00/1.40	3.00/3.50	0.08/0.15	—
Bower 315**	Low	0.10/0.15	—	—	1.5	3.0	5.0	—
Timken CBS600	Low	0.16/0.22	0.50/0.70	0.90/1.25	0.25/1.65	—	0.90/1.10	—
Timken CBS1000	Low	0.18/0.23	0.40/0.60	0.40/0.60	0.90/1.20	—	4.75/5.25	Vanadium 0.75/1.00

\* Not AISI Standard Steel

\*\* Typical chemistry from engineering alloys

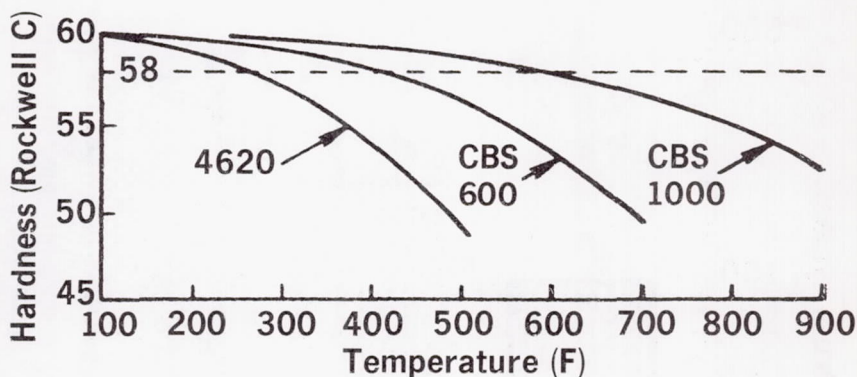


FIGURE 2.—Hot hardness of three carburized bearing steels.

M-50 materials for equivalent material hardnesses. Therefore, a hot hardness criteria for bearing material selection appears to be reasonable for the more common bearing alloys.

Fatigue and endurance studies were reported in references 5, 12, and 13 at 600° F with 120-mm bore angular-contact ball bearings. Lives obtained with bearings made of AISI M-50 material having a nominal room temperature Rockwell C hardness of 63 were approximately 12 times that calculated using the Lundberg-Palmgren calculations. Life values calculated from the Lundberg-Palmgren relations will be referred to as AFBMA or catalog life. For the same size bearing made from WB-49 material having a nominal room temperature Rockwell C hardness of 65, fatigue lives of only 4 percent of that of the M-50 bearings were obtained. For these tests, the lubricant was a synthetic paraffinic oil having an anti-wear additive under a low oxygen environment. Material hardness, heat treatment, heat of material, and lubricant batch were controlled variables in addition to the test conditions.

Endurance tests (ref. 4) with 25-mm bore angular-contact ball bearings indicated that at 600° F, AISI M-1 material of nominal Rockwell C hardness 63 may be as good as the AISI M-50 material of an equivalent room-temperature hardness. These tests were also run with a synthetic paraffinic oil having an anti-wear additive under a low oxygen environment.

Criteria for selecting a material, in addition to fatigue life and operating temperature, include fabrication costs. Generally bearings made from the M-series materials can cost as much as 50 percent more to manufacture than if made from standard SAE 52100 because of grinding difficulties. At temperatures below 350° F, there appears to be no technical or cost advantage of conventional M-series materials over SAE 52100.

No data are available in the open literature which compare the fatigue



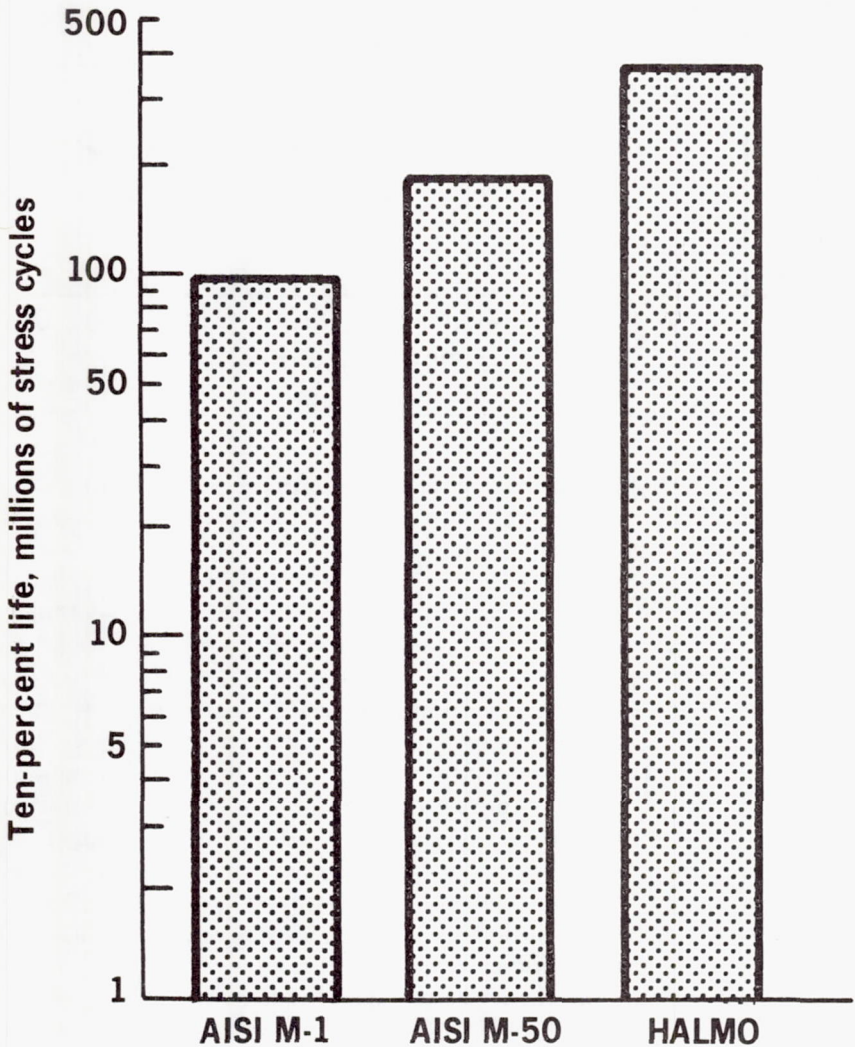


FIGURE 3.—Comparative lives of three bearing materials in NASA Spin rig. Specimens,  $\frac{9}{16}$ -in. diameter balls; maximum Hertz stress, 800 000 psi; temperature, ambient (no heat added); Rockwell C hardness, 62; lubricant, diester.

lives of carburized and through-hardened materials. On the basis of room-temperature component hardness, through-hardened materials may give the higher dynamic capacity. However, where shock and high vibrational loads are present, carburized materials may be less susceptible to catastrophic failure because of the soft, ductile inner core.

## PROCESSING VARIABLES

## Controlled Fiber

Races and rolling elements of most bearings are formed by **forging**, and thus possess a fiber pattern that reflects the flow of metal during the forming. Bearings manufactured by two methods are shown in figure 4. One method produced races with conventional fiber orientation; the other was forged. Fatigue data shows an increase in life for bearings having side grain or fiber flow parallel to the race compared to those with the end-grain races or fiber nearly perpendicular to the running track (refs. 6 and 14). The life of parts seems less sensitive to steel quality variations when forged with the fiber flow parallel to the race.

Standard bearings are commercially available with controlled-fiber orientation. These can have dynamic capacities as much as twice those of more conventionally manufactured bearings.

## Ausformed Bearings

Ausforming consists of hot rolling while the steel is in a metastable austenitic condition. To apply this method to bearing steels, they must have a sluggish transformation behavior.

Bearing tests indicate that bearings of ausformed AISI M-50 are superior to those of normally processed M-50 (refs. 15 and 16). Also, a

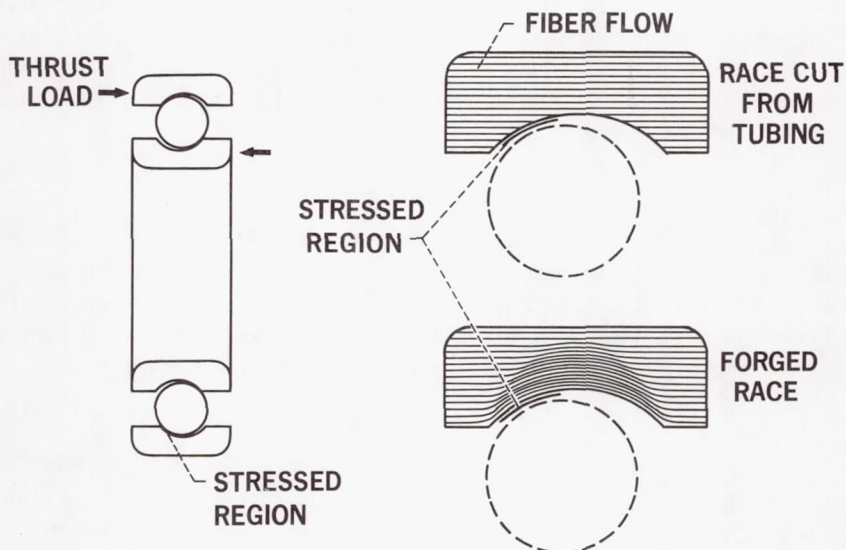


FIGURE 4.—Fiber flow in bearing races.

relation exists between the amount of deformation during ausforming and fatigue life (ref. 15).

Bearings manufactured in accordance with the ausformed process are commercially available on a limited basis at a premium cost. They can be specified for applications where rolling-element fatigue is a critical problem and extreme reliability is desired. The bearings generally have a dynamic load capacity twice that of conventional ones.

#### Component Hardness

In general, increased dynamic capacity, and thus bearing fatigue life, can be obtained with increased bearing component hardness in the hardness range where the material maintains its ductility. No correlation exists between bearing life and mechanical strength properties such as elastic limit, yield strength, and ultimate strength (refs. 11 and 17).

Results of rolling-element fatigue tests (refs. 18 and 19) performed on SAE 52100, 207-size, deep-groove, ball bearings grouped according to difference between hardness of the rolling elements in the bearing and the hardness of the inner race ( $\Delta H$ ) are shown in figure 5. These bearings exhibited a maximum life for a  $\Delta H$  of approximately 1 to 2 points Rockwell C.

Generally in a rolling-element system, the rolling element receiving the greatest number of stress cycles per unit time should be 1 to 2 points Rockwell C softer than mating elements. Current heat-treating procedures result in bearing components with significant hardness variations. Consequently where care is not taken to properly match balls and races, no improvement in fatigue life can be expected by specifying  $\Delta H$ , as is illustrated in table 3. Bearings with controlled  $\Delta H$  can be specially ordered, although a premium price will be paid for matching the rolling elements to the races.

#### Residual Stress

Among the factors that can affect the fatigue life of a rolling element is subsurface residual stress. Residual stresses can be produced in a material by such methods as heat treating, rolling, peening, and severe grinding. Residual stress can either increase or decrease the maximum shearing stress (ref. 20) according to the following equation:

$$(\tau_{\max})_r = \tau_{\max} - \frac{1}{2}(\pm S_r)$$

where  $(\tau_{\max})_r$  is the maximum shearing stress modified by the residual stress,  $S_r$ , and  $\tau_{\max}$  is the maximum shearing stress due to the applied load. The positive or negative sign indicates a tensile or compressive residual stress, respectively. A compressive residual stress would reduce the magnitude of the maximum shearing stress (because the value of



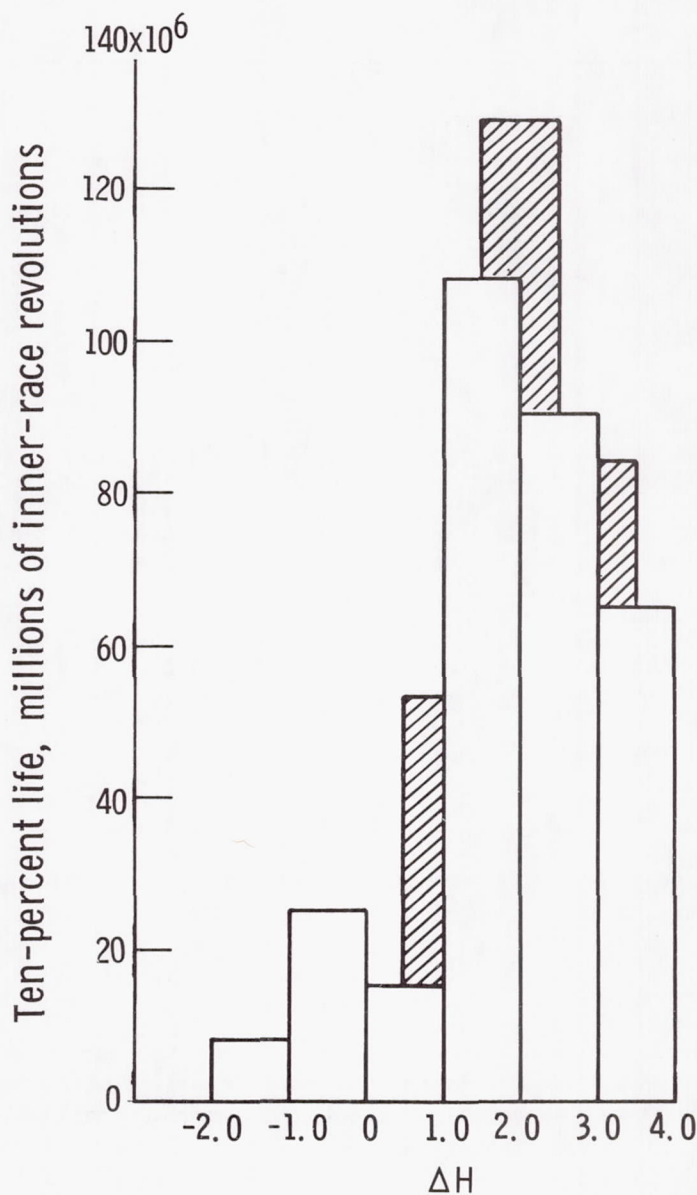


FIGURE 5.—Ten-percent life as function of  $\Delta H$  (difference in Rockwell C hardness between balls and races) for 207-size, deep-groove, ball bearings. Material: SAE 52100 steel balls and races; radial load, 1320 lb; speed, 2750 rpm; lubricant, super-refined naphthenic mineral oil; no heat added.

TABLE 3.—*Experimental Effect of Differential Hardness and Hardness Tolerances on Fatigue Life of 207-Size SAE 52100 Deep-Groove Ball Bearings*

Hardness tolerances	Specified $\Delta H$				
	0	1	2	3	
Races $\pm 0$	$-\frac{1}{2}$ to $\frac{1}{2}$	$\frac{1}{2}$ to $1\frac{1}{2}$	$1\frac{1}{2}$ to $2\frac{1}{2}$	$2\frac{1}{2}$ to $3\frac{1}{2}$	$\Delta H$ range
Rolling elements $\pm \frac{1}{2}$	$\sim 0.07$	0.34	1.00	0.55	Relative life
	-1	1.1	1.7	1.3	Weibull slope
Races $\pm \frac{1}{2}$	-1 to 1	0 to 2	1 to 3	2 to 4	$\Delta H$ range
Rolling elements $\pm \frac{1}{2}$	0.20	0.52	0.66	0.62	Relative life
	1.1	1.1	1.5	1.4	Weibull slope
Races $\pm 1$	$-1\frac{1}{2}$ to $1\frac{1}{2}$	$-\frac{1}{2}$ to $2\frac{1}{2}$	$\frac{1}{2}$ to $3\frac{1}{2}$	$1\frac{1}{2}$ to $4\frac{1}{2}$	$\Delta H$ range
Rolling elements $\pm \frac{1}{2}$	0.16	0.49	0.62	0.71	Relative life
	0.9	1.2	1.4	1.4	Weibull slope
Races $\pm 1$	-2 to 2	-1 to 3	0 to 4	1 to 5	$\Delta H$ range
Rolling elements $\pm 1$	0.20	0.50	0.60	0.66	Relative life
	0.6	1.3	1.4	1.4	Weibull slope
Races $\pm 2$	-3 to 3	-2 to 4	-1 to 5	0 to 6	$\Delta H$ range
Rolling elements $\pm 1$	0.27	0.32	0.53	0.60	Relative life
	1.0	1.0	1.3	1.4	Weibull slope

$\tau_{\max}$  is negative) and increase fatigue life according to the inverse relationship between life and stress where

$$\text{Life} \propto \left[ \frac{1}{(\tau_{\max})^r} \right]^9$$

Research reported in references 21 and 22 indicates that these compressive residual stresses do indeed increase fatigue life as illustrated in figure 6. There is shown the stress pattern for raceways with essentially no residual stresses (termed Group A) and raceways wherein there were induced compressive residual stresses greater than 140 000 psi (termed Group B). Tests were conducted on 50 bearings from each of Groups A and B. Bearing Group B, having the residual stresses, exhibited twice the life of Group A having no residual stresses. For lightly and medium-loaded bearings, residual stress can increase life markedly. However, for heavy loads the residual-stress effect is not very significant.

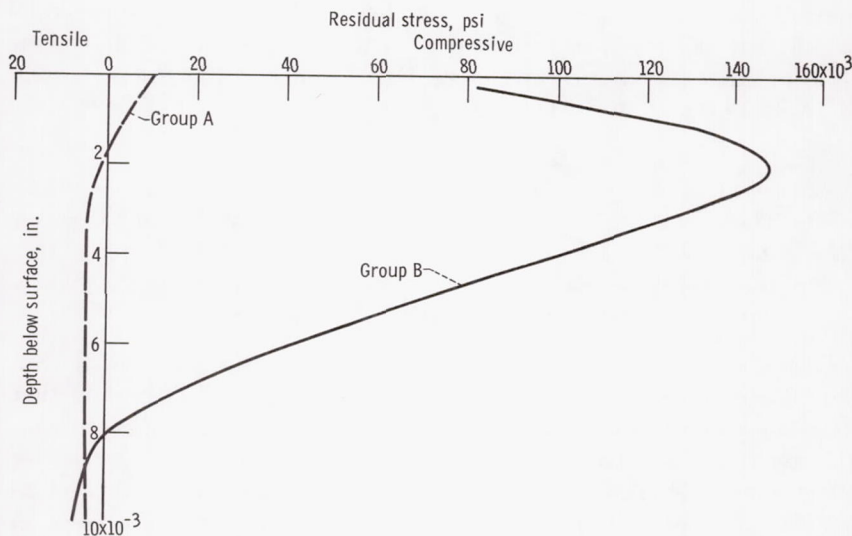


FIGURE 6.—Tangential residual stress patterns for two groups of 40-mm bore-size bearing races.

Almen (ref. 23) was the first to study residual stresses in rolling-element bearings in 1944. From these initial studies it was concluded that residual stresses increased in a rolling-element bearing with running time. Later results have substantiated Almen's findings (refs. 24 and 25). In addition, research reported in reference 26 indicated that the compressive residual stresses induced in a bearing inner race during running is a function of the hardness of the rolling elements minus the race hardness, termed  $\Delta H$ . An interrelation was reported between  $\Delta H$ , induced compressive residual stress, and fatigue life. Further, a qualitative correlation was reported between relative predicted lives based upon measured residual stress and relative lives determined experimentally.

#### MATERIAL CLEANLINESS

Research performed by Bear, Butler, Carter, and Anderson (refs. 14, 27, and 28) substantiated the early findings of Jones (ref. 29) that one mode of rolling-element fatigue is due to nonmetallic inclusions. These inclusions act as stress raisers similar to notches in tension and compression specimens or in rotating beam specimens. Incipient cracks emanate from these inclusions, enlarge and propagate under repeated stresses forming a network of cracks which form into a fatigue spall or pit. In general the cracks propagate below the rolling-contact surface approximately  $45^\circ$  to the normal; i.e., they appear to be in the plane of



maximum shearing stress. Carter (ref. 14) made a qualitative generalization that the location of an inclusion with respect to the maximum shearing stress is of prime importance. Based on observations of inclusions in SAE 52100 and AISI M-1 steels, Carter concluded that

- (1) Inclusion location is of primary importance
- (2) Size and orientation are also important
- (3) The oxides and larger carbides are more harmful than the softer sulfide inclusions, and
- (4) Inclusions, carbides, and irregular matrix conditions appear slightly less harmful to fatigue life in SAE 52100 than in AISI M-1.

Carter's conclusions were substantiated by Johnson and Sewell (ref. 30). The results of their work are summarized in figure 7. They show that as the total number of alumina and silicates increase, fatigue life decreases. However, they indicate that the increase in sulfides may have a positive effect upon fatigue life. In addition to inclusions, material defects such as microcracks, trace elements, or unusual carbide formations present in the material can contribute to failure.

Techniques have been developed by the bearing companies and engine manufacturers to nondestructively inspect bearings in the subsurface region where classical fatigue initiates from these inclusions and material defects. Among the methods used are magnaflux, ultrasonics, mechanical hysteresis, eddy currents, and magnetic permeability.

Koved and Rospond (ref. 31) have reported the most successful use of ultrasonics as an inspection tool. Roller bearings made of AISI 8620 were arranged into three groups based on differences in the size and the frequency of ultrasonic indications. The individual bearings were rated by a method that assigned a numerical value to their ultrasonic pattern. The number was based on the frequency of indications and included a

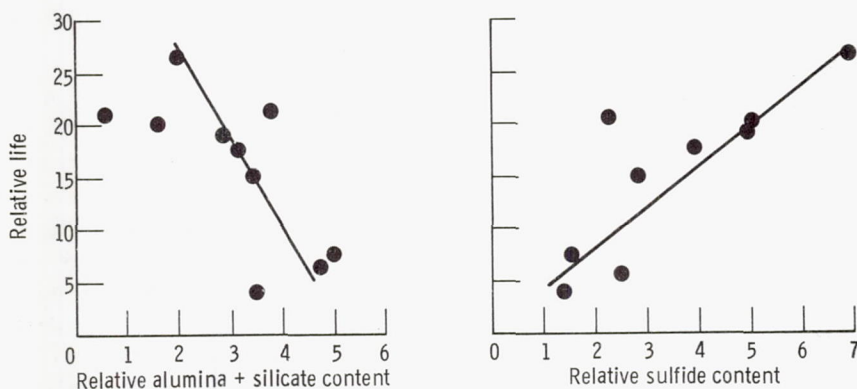


FIGURE 7.—Relationship between life and inclusion content.

weighting factor for size. These numerical ratings were subsequently converted into qualitative ratings.

The fatigue life distributions for the ultrasonically-rated groups of bearings are shown in figure 8. The three performance curves separate nicely in the same order as their ultrasonic classification. The group characterized as ultrasonically poor demonstrated the worst performance, average quality was intermediate, and good quality clearly exhibited the best fatigue-life distribution. Thus ultrasonic inspection has the capability of differentiating, on a statistical or group basis, relative material quality in terms of life performance.

#### RETAINERS OR CAGES

The retainer or cage maintains the proper distance between the rolling elements. In conventional rolling-element bearings, both metallic and nonmetallic retainers are used. Under normal temperatures, almost all roller bearings and a large percentage of ball bearings are equipped with stamped retainers of low-carbon steel or machined retainers of either iron-silicon bronze or a lead brass. Precision bearings are usually equipped

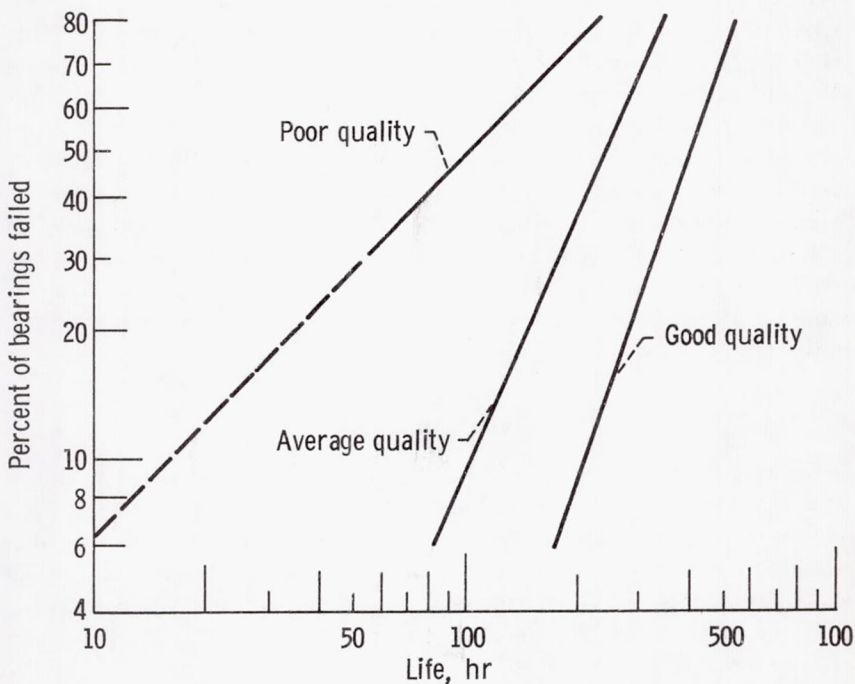


FIGURE 8.—Fatigue life distribution of 309-size roller bearings made from AISI 8620 steel grouped into three quality ratings based on ultrasonic response.

with machined rather than stamped retainers, and these are copper alloys. In some applications, where marginal lubrication exists during operation, silver plating is used on the bronze (ref. 1).

For most applications, the retainer can be inner-race or outer-race riding—that is, one of the races guides the retainer. For high-speed applications it is preferable to have a one-piece, machined, outer-race-riding retainer of the thin-line type as seen in figure 9. This type of cage allows a greater amount of lubricant into the bearing for cooling purposes than the conventional type. Also, the retainer surface is easy to lubricate because centrifugal force delivers the lubricant to the contacting surfaces of the retainer. The outer-race-riding retainer also wears itself into balance. On the other hand, the problem of dirt in the lubrication system often results in severe race wear with an outer-race-riding retainer.

Retainers are a much more severe problem in small-bore bearings than in larger ones. In extreme-speed applications of small-bore bearings, it is frequently necessary to use a silver-plated, semihard, tool-steel retainer rather than a bronze, and oil-mist lubrication rather than recirculating oil to reduce churning losses. In large-bore bearings, retainer failure is much less common than in small-bore bearings. The retainer should be as light as possible consistent with required strength and should offer minimum resistance to lubricant flow in and out of the bearing.

Relative wear of six high-temperature cage materials is given in figure 10 (ref. 32). S-Monel, M-1, 440C stainless steel and a polyimide polymer showed low potential wear at 500° F.

For the alloys in figure 10, retainer wear usually decreases with increased material hardness. Thus, cages for high-temperature applications should be heat-treated to maximum practical hardness.

Besides the polyimide polymer shown in figure 10, which is limited to temperatures under 700° F, other nonmetallic retainer materials are commonly used. Among the most common is phenolic. Operating temperature for phenolics is limited to approximately 270° F. PTFE fluorocarbon is also a widely used retainer material and is limited to about 450° F.

#### CERAMICS AND CERMETS

Aerospace technology dictates a need for bearings to operate reliably at temperatures to 2000° F. Since the temperature range between 800° and 2000° F is beyond the range in which current ferrous and nonferrous bearing materials are capable of operating, the more refractory materials and compounds must be considered. Among these materials are alumina, titanium carbide cermets, and silicon carbides. A relatively large amount of research and development has been performed with alumina and titanium carbide cermets in rolling-element bearings. Data have also been reported on the use of silicon carbide for this purpose.



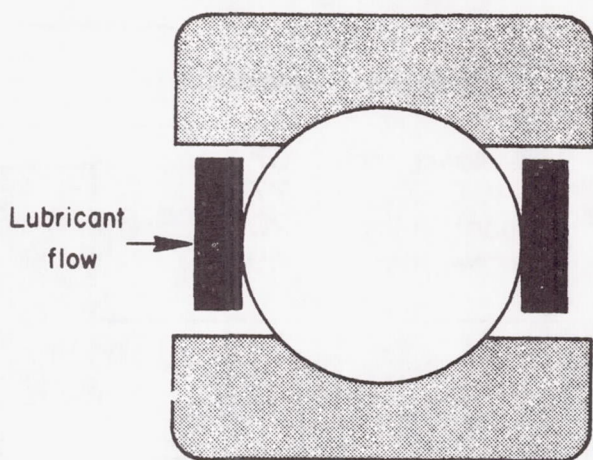
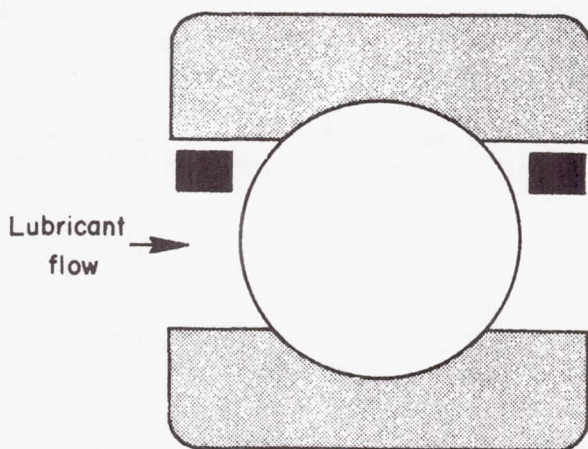
**Conventional Retainer****Thinline Retainer**

FIGURE 9.—Thinline, one-piece, ball bearing retainer does not impede lubricant flow.

#### **Alumina**

The properties of alumina are given in table 4. Data reported in references 33 and 34 indicate that, in both sliding and rolling contact at temperatures to 1200° F, alumina exhibits friction and wear characteristics somewhat similar to those of conventional bearing steels over the

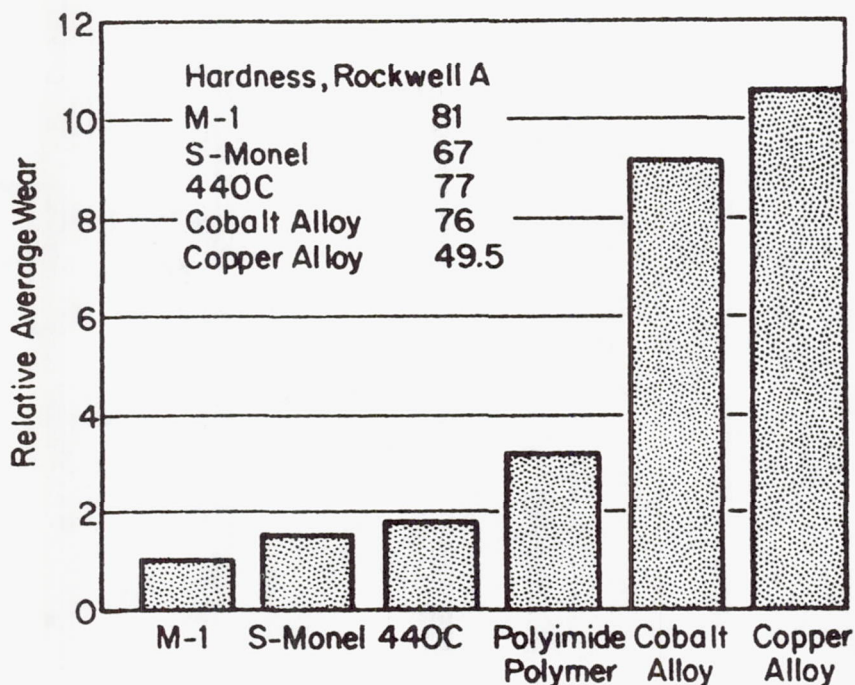


FIGURE 10.—Effect of various materials on cage wear in an inert environment. Test conditions: ambient temperature, 500° F; shaft speed, 1200 rpm; load, 100 lb; duration, 30 min; lubricant, super-refined naphthenic mineral oil.

same temperature range. Additionally the coefficient of friction of alumina sliding unlubricated on several materials at temperatures up to 1600° F is comparable with that of M-2 steel sliding unlubricated on the same materials at temperatures to 1000° F (ref. 33).

Fused alumina balls (96.0 percent  $\text{Al}_2\text{O}_3$ ) were run unlubricated under oscillatory motion on several plate materials at temperatures from 600° to 1200° F and at maximum Hertz stresses from 637 000 to 1 012 000 psi (ref. 34). The wear in all tests was light, but in some tests the balls fractured. However, fracture at such severe stresses would not be unexpected.

Surface failure tests in rolling contact were conducted in the five-ball fatigue tester with both hot-pressed and cold-pressed alumina balls (refs. 35 and 36). The hot-pressed alumina balls were dark gray and exhibited some randomly distributed darker streaks and spots. The major impurities present in the 99 percent-pure alumina were magnesium, silicon, and iron, which produced a complex spinel phase with a lattice parameter similar to that of nickel aluminate. This material contained about 0.6 volume

TABLE 4.—Average Physical Property Data of Refractory Materials

Property	Temp. (°F)	Alumina	Self-bonded silicon carbide	Nickel-bonded titanium carbide cermet
Compressive strength (psi)	70 1200 2012	427 000 85 000	150 000	450 000 296 000
Modulus of elasticity (psi)	70 1600 2012 2200	$52.4 \times 10^6$ $\sim 41 \times 10^6$	$69 \times 10^6$ $49 \times 10^6$	$57 \times 10^6$ $48 \times 10^6$
Poisson's ratio	70	0.26	0.183	0.20
Hardness	70 1400	2000 <sup>a</sup>	2740 <sup>a</sup>	89 <sup>b</sup> 74 <sup>b</sup>
Melting point (°F)		3659–3723	4680 (sublimes at $T > 3600^\circ \text{F}$ )	TiC:5700 Ni:2647
Dynamic load capacity (percent) AISI M-1 steel 100%		1 <sup>c</sup> –7 <sup>d</sup>	1	3

<sup>a</sup> Knoop 100 gm scale<sup>b</sup> Rockwell A scale<sup>c</sup> Cold-pressed and sintered<sup>d</sup> Hot-pressed

percent pores; its surface finish was 0.3- to 0.5- $\mu$ in. rms. The dynamic load carrying capacity at room temperature was found to be 7 percent that of AISI M-1. At 2000° F the dynamic load capacity of alumina was approximately 2 percent that of AISI M-1 at room temperature.

The balls made of the 99 percent-pure, cold-pressed-and-sintered alumina material ranged in color from white to cream because of small amounts of magnesium, calcium, and silicon impurities. On the basis of color alone, the balls, all from the same batch, were not found to be of uniform composition. Further examination of these balls indicated that the surface of the as-received specimens showed a roughness of from 3- to 8- $\mu$ in. rms. This relatively poor surface was the result of 4.3 volume percent pores in the material. The cold-pressed alumina exhibited dynamic load capacity of 15 percent that of the hot-pressed alumina or approximately 1 percent that of AISI M-1 bearing steel (refs. 35 and 36).

#### Silicon Carbide

As indicated in table 4, self-bonded silicon carbide contains approximately 12 volume percent of silicon which is relatively soft and has a low modulus of elasticity. Unlubricated rolling-element tests of this



material (ref. 37) indicated minimal wear relative to other materials reported in reference 37 at a maximum Hertz stress of 375 000 psi for over 250 million stress cycles of operation. Macroscopic examination of the running track revealed no surface cracking or spalling. Additional research reported in reference 38 indicated that a titanium carbide-silicon carbide combination resulted in approximately the same sliding friction coefficients as that of a titanium carbide-alumina combination.

At 80° F in the five-ball fatigue tester this material had a dynamic load capacity of approximately 1 percent that of AISI M-1 steel (refs. 36 and 39). Preliminary tests with molybdenum disulfide-argon mist lubrication at a maximum Hertz stress of 400 000 psi show that self-bonded silicon carbide exhibits excessive plastic deformation at temperatures as low as 1600° F. This would indicate that this material is limited to less severe conditions of temperature and stress.

#### Titanium Carbide

In reference 40 the data indicated that a 20-mm bore, titanium carbide cermet ball bearing was capable of running to temperatures of 1200° F for periods of 2 to 3 hours at DN\* values of approximately  $\frac{1}{2}$  million with a solid-film lubricant. Typical properties of a titanium-carbide cermet are given in table 4.

---

\* DN is defined as the product of the bearing bore in mm and bearing speed in rpm.

Research reported in reference 41 indicated that titanium carbide cermet exhibited friction coefficients comparable with those of alumina sliding unlubricated on similar materials to temperatures to 1650° F. Also, development work was reported in reference 42 with a bearing with titanium carbide races and alumina balls run at moderate loads with molybdenum disulfide lubricant carried in an inert gas. The bearings operated at temperatures to 1500° F for periods as long as 8 hours. Failure of these bearings was by pitting of the titanium carbide races. However, tests performed in the five-ball fatigue tester (refs. 36 and 39) with a nickel-bonded titanium carbide cermet at a maximum Hertz stress of 310 000 psi at temperatures as low as 1100° F indicated excessive plastic deformation of the material. This result would tend to indicate that the nickel-bonded titanium carbide is limited to a less severe condition of temperature and stress.

The surface failures in titanium carbide very closely resembled those in alumina. Because of the lack of porosity, the relatively small number of surface pores, and the presence of the ductile matrix, the failure mechanism in this material would not be expected to be the same as that in alumina (i.e., brittle fracture). (The nickel-titanium carbide cermet

involves a ductile nickel matrix in which 71 volume percent of fine, evenly divided carbide particles are uniformly dispersed.) Brittle fracture may have occurred, however, in the carbide particles because of the high volume fraction of this phase present. The dynamic load capacity of this material at room temperature was approximately 3 percent of AISI M-1 steel.

#### Conclusions

The ability of a ceramic or cermet material to be functional in rolling-contact applications appears to be related to its physical properties and surface condition. Consequently surface finish appears to be an important criterion for long life. Since surface finish was related to the amount of porosity or to the presence of a weak second phase, such conditions should be avoided for proposed high-temperature bearing materials.

On the basis of the investigation of Parker et al. (refs. 35, 36, and 39), present-day ceramic fabrication techniques appear to be unable to supply high-quality material for rolling elements for high-temperature applications. As ceramic materials approach homogeneity and zero porosity, however, high-temperature bearing reliability and load capacity should improve.

#### MATERIAL DISLOCATIONS

Fatigue failure of rolling-element bearings continues to be the limiting factor of bearing life in spite of recent materials and lubricant improvements. Further improvements in fatigue life will probably result only from a basic understanding of the fatigue mechanism.

Limited plastic deformation is assumed to be a prerequisite for rolling-element fatigue. Plastic deformation involves generation and motion of dislocations. Interaction of the moving dislocations results in work-hardening and crack initiation. It is presumed that these fundamental concepts are involved in the process of rolling-element fatigue. General dislocation theory should provide a powerful tool for the development of a deeper understanding of the rolling-element fatigue process.

The deformation characteristics of all solids result from the generation, motion, and interaction of dislocations. As the complexity of the solid increases with grain boundaries, multiple phases, and impurity precipitates, the understanding of deformation becomes increasingly difficult. A simple solid with well-established slip systems is necessary as a beginning material to study basic rolling-contact deformation and fatigue mechanisms.

There are some significant differences between rolling-contact stress conditions and general structural stress conditions that must be kept in mind when analyzing the interplay between dislocations and applied stress. For the general case of structural stress, a member is subjected to a relatively static stress field, and, as in the case of a beam under tensile



load, the entire section is stressed. Dislocations in the part are acted on by the stress field and move until they reach the surface or are stopped by other dislocations or contaminants. In rolling contact, a very small zone in a body is acted on by a moving concentrated stress field. Therefore, instead of moving until stopped by some obstacle, the dislocations may move only briefly as they are acted upon by the stress field when it moves by. In structural stress conditions, large arrays of dislocations are moved simultaneously, and many local effects are averaged out. In rolling contact, local effects and individual dislocation interactions become significant because of the limited time and field of dislocation-stress interplay.

Early experiments (ref. 43) showed single-crystal, high-purity magnesium oxide to fulfill the requirements of a simple model bearing material. It has well-established slip systems, and experimental convenience permits simple analysis of deformation from rolling-contact stresses as shown in figure 11. For this material the operating slip systems were found to be those associated with the maximum applied shearing stress (ref. 43). Dislocations propagated from sources located beneath the surface near the region of the calculated maximum shear stress. The depth of dislocation penetration was found to be greater on planes of maximum shear stress that were parallel to the rolling direction than on similar planes that were inclined at 45 degrees to the rolling direction (fig. 12). As the number of reciprocating rolling cycles increased, the track width increased, reached a plateau, and then increased again. Spalling, very similar to fatigue spalls seen in actual bearing elements, often occurred during the second increase of the track width.

Subsequent studies (ref. 44) revealed several other features. The depth of dislocation propagation decreased with increasing rolling velocities. This could be explained in terms of dislocation dynamics. However, variations in track width with rolling-contact cycles could not be explained. Rolling with a lubricant present under boundary lubrication conditions caused a change in slip mode that led to spalling in  $10^3$  cycles as opposed to at least  $10^5$  cycles for dry contact. It was speculated that the lubricant chemically interacted with the magnesium oxide surface to cause this effect.

It was reported in reference 45 that hardness measurements on magnesium oxide subjected to reciprocating rolling contact show the same variation with the number of rolling-contact cycles as track width. It was therefore concluded that track-width measurements were a measure of dislocation density and extent of interaction in the same manner as hardness.

This research (ref. 44) showed that the formation of dislocations and their associated density and interaction was significantly different under EHD conditions as compared to unlubricated or boundary lubricated



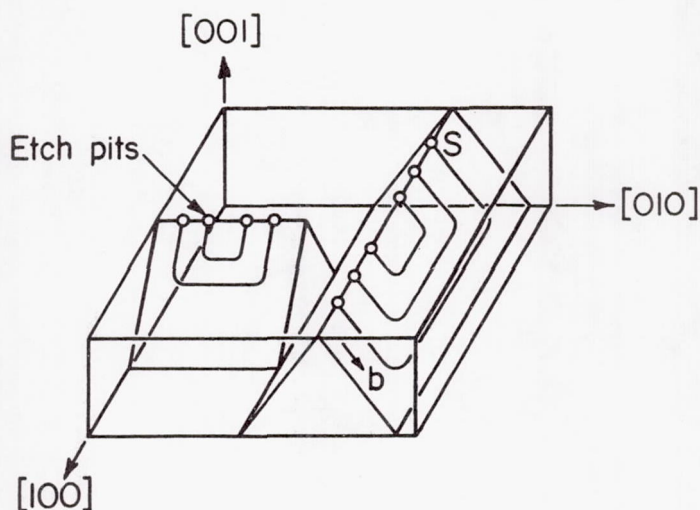
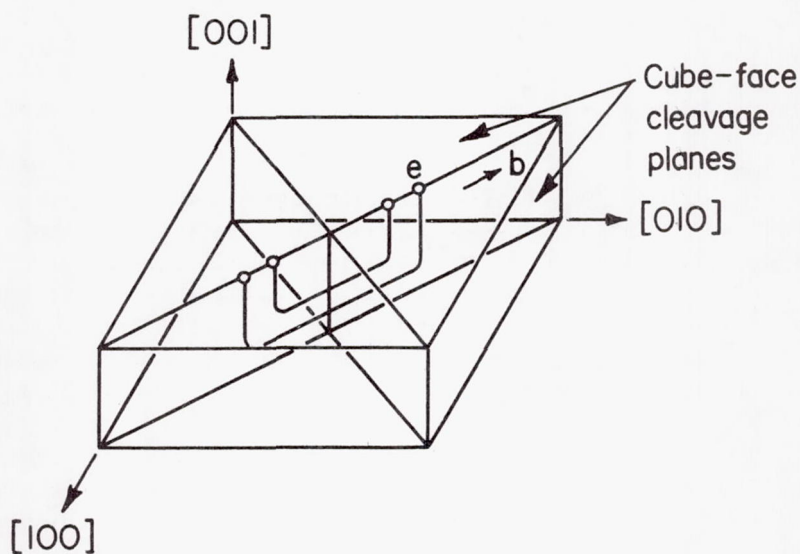
a.  $\{110\}_{45^\circ}$  Slip Planesb.  $\{110\}_{90^\circ}$  Slip Planes

FIGURE 11.—Crystallographic direction and slip planes in magnesium oxide.

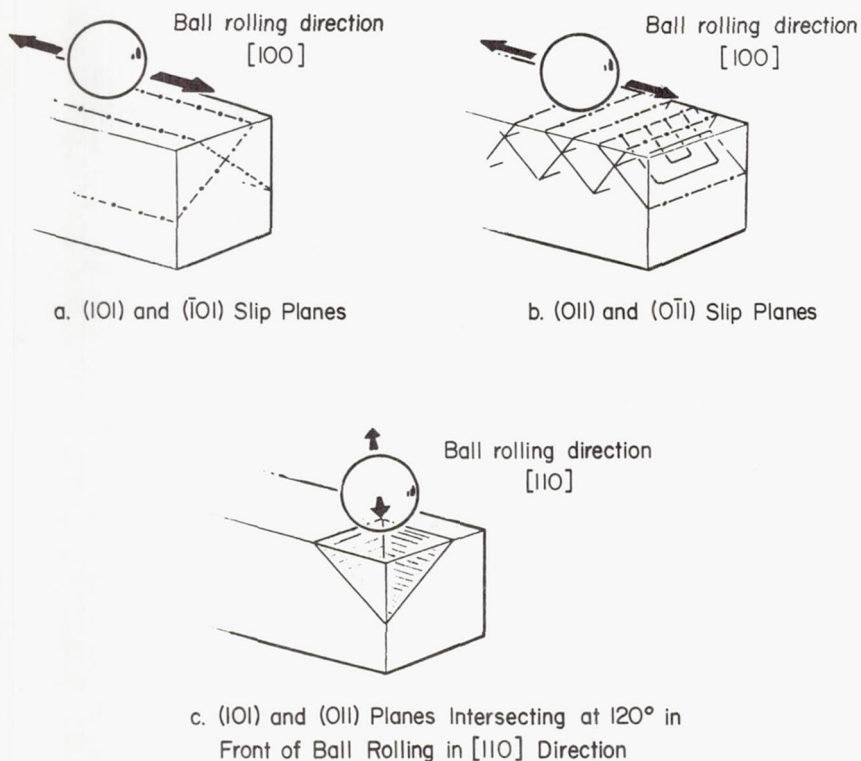


FIGURE 12.—Preferred slip systems in single-crystal MgO for various ball rolling directions.

conditions. Rolling under EHD conditions caused slip to occur in narrow bands across the track. As rolling continued, the bands widened and eventually merged to form a uniform dislocation configuration across the track. In addition, the track width and hardness increased to maximum levels more rapidly under EHD rolling conditions. The track width after  $10^6$  rolling-contact cycles was twice that with low-speed-reciprocating rolling, but the track hardness was the same for both conditions of rolling.

With an understanding of the basic rolling-contact deformation and fatigue mechanisms in MgO, similar studies can be performed on more complicated materials that have additional slip systems, second phases, and grain boundaries. Materials chosen for such studies require a sufficiently high yield strength to withstand high rolling-contact stresses without excessive plastic deformation. The usefulness of etch-pit techniques in analyzing dislocation configurations resulting from rolling-contact deformation makes established etch-pit techniques an additional

desirable feature. Possible materials include silicon-iron and semiconductor single crystals and polycrystals. When the influence of additional factors of multiple slip systems, grain boundaries, surface activity, and second-phase particles is determined, the probability of analyzing the basic fatigue mechanism in actual bearing materials will be increased.

### CONCLUSIONS

(1) Criteria for selecting a bearing material, in addition to fatigue life and operating temperature, include fabrication costs. Generally bearings made from the M-series materials can cost as much as 50 percent more to manufacture than if made from standard SAE 52100 because of grinding difficulties. At temperatures below 350° F, there appears to be no technical or cost advantage of conventional M-series materials over SAE 52100.

(2) No data are available in the open literature which compare the fatigue lives of carburized and through-hardened materials. On the basis of room-temperature component hardness, through-hardened materials may give the higher dynamic capacity. However, where shock and high vibrational loads are present, carburized materials may be less susceptible to catastrophic failure because of the soft, ductile inner core.

(3) Angular-contact bearings having races with fiber orientation parallel to the ball-race contact can have dynamic capacities as much as twice those of more conventionally manufactured bearings.

(4) Bearings made by ausforming can have a dynamic load capacity twice that of conventional ones.

(5) In a rolling-element system using SAE 52100 steel, the rolling element receiving the greatest number of stress cycles per unit time should be 1 to 2 points Rockwell C softer than mating elements in order to achieve maximum fatigue life.

(6) Compressive residual stress can decrease maximum shearing stress from operating loads and thus can increase fatigue life. The amount of increase depends on the induced residual stress and the applied load.

(7) Techniques have been developed to nondestructively inspect bearings in the subsurface region where classical fatigue initiates from inclusions and material defects. Among the methods used are magnaflux, ultrasonics, mechanical hysteresis, eddy currents, and magnetic permeability.

(8) Bearing retainers or cages should be as light as possible consistent with required strength and should offer minimum resistance to lubricant flow in and out of the bearing. Cages for high-temperature operation should be heat-treated to maximum practical hardness.

(9) In general, ceramic and cermet rolling elements have dynamic load capacities approximately 1 to 7 percent that of conventional bearing steels.

(10) Single-crystal magnesium oxide under rolling-contact stress con-



ditions exhibited spalling similar to conventional rolling-element materials. The spalling in the magnesium oxide can be attributed to dislocation density and interaction. This dislocation density and interaction was significantly different under elastohydrodynamic conditions as compared to unlubricated or boundary lubricated conditions.

#### DISCUSSIONS

**S. Allen (MIT Instrumentation Laboratory, Cambridge, Massachusetts)**

The authors state that the maximum obtainable hardness of 440C stainless steel is approximately  $R_{c60}$ . In their view, 440C therefore compares unfavorably with the other bearing steels presented in their figure 1 since, even at room temperature, it is so near the somewhat arbitrary lower limit of  $R_{c58}$ . In 1963 we reviewed 440C heat treatments used by instrument ball bearing manufacturers and found  $R_{c60}$  and load derating to be the general rule. Experimentally it was found that, by increasing the degree of austenization, oil quenching, subcooling (usually to liquid nitrogen temperature), and double tempering, 440C hardness averaged  $R_{c63}$ . Oil quenching adds to both hardness and corrosion resistance, and subcooling is necessary to convert the otherwise excessive retained austenite. Mechanical properties of approximately 25 production heats of 440C for instrument bearings thus heat-treated were very consistent with only one appreciably low ( $R_{c62}$ ) in hardness. Considering the higher room-temperature hardness and 440C resistance to softening by tempering, the maximum operating temperature may well be higher than the listed 300° F. Compositional variants of 440C, such as the 14 Cr-4 Mo cited by Bamberger\* can furnish higher initial hardness and greater tempering resistance. Aside from corrosive end-use environments, stainless steel instrument bearings are often specified because this prevents most pre-use corrosion. Load derating of 440C compared to 52100 is currently much less common.

440C was reputed to be difficult to grind and finish, perhaps analogous to the M-series mentioned. The most closely monitored bearing manufacture showed an initial adaptation phase and no subsequent difficulties. A later, limited reintroduction of 52100 surprisingly required adaptive development, indicating that the critical factor was not the specific material but the change from that to which the manufacturing processes had been "tuned."

Storage and controlled supply of lubricant is an important function of the retainer in bearings, such as instrument bearings, where a continuous external supply is not provided and where, for low and/or constant torque or other reasons, the lubricant quantity must be stringently

---

\* See Bamberger's lecture on Effect of Materials—Metallurgy Viewpoint.

limited. In these applications, which were not included among the authors' review, surface and volume porosity are important properties. Normal cloth-based phenolic retainers are limited since they exhibit only surface storage of lubricant. On the other hand, larger quantities for circulation (ref. 46) can be provided by utilizing the interconnected volume porosity of sintered compacts, e.g., nylon (ref. 47) or metal (ref. 48), or special composites. Currently instrument bearings usually have one-piece retainers of cloth-based phenolic or sintered nylons. Research aimed at eliminating retainers in some instrument applications encompasses other methods of providing a lubricant supply, as by porous balls separating the normal ones.

Matrix steels are compositional modifications intended to eliminate the primary carbides found in the standard steel grades. They have been developed and tested on premises partly resembling those listed for ausforming, i.e., potential benefits of less, smaller, and more uniform carbides. Among these hoped-for improvements is the one (not-listed) of smoother finishes that might be achieved when the material hardness is more nearly constant on a microscopic scale. This benefit was not found especially significant in gyro bearings (ref. 49). Concomitantly an  $R_{c1-2}$  decrease from hardness of the standard composition was noted (ref. 49). Subsequent commercial availability of matrix alloys (ref. 50) makes some further exploration more attractive since some benefits might accrue without the special, expensive manufacturing efforts involved in ausforming or other thermomechanical treatments (ref. 51).

#### LECTURERS' CLOSURE

The authors would like to thank Dr. Allen for his discussion. Dr. Allen, who has performed research in the field of instrument bearings, supplements our paper by his comments in this area.

Regarding the 440C stainless steel material, commercial heat treatments commonly available at the time of writing our paper allowed for maximum Rockwell C hardnesses of only 60. The authors subsequently performed hardness tests on samples of 440C stainless steel heat-treated by Dr. Allen's method. The measured nominal hardness of the samples was Rockwell C 62. As a result, the authors agree that the upper temperature limit for the 440C material stated by the authors may be higher. Additional data of hardness as a function of temperature for this material should be generated.

The subject of surface finish is an interesting one from the standpoint of lubrication mode. Where the surfaces are not completely separated by an elastohydrodynamic film, asperity interaction can cause high torque values and torque fluctuation. Improved surface finish decreases the probability of surface interaction and torque fluctuation. Bearing manufacturers recently have developed honing methods to finish bearing races



for materials such as conventional SAE 52100 and the M-series materials, whereby the resultant surface finish is approximately 1- $\mu$ in. rms. Surface finishes of 2- to 3- $\mu$ in. rms can be obtained without the improved honing methods. Today, most commercial bearings of reasonably good quality have surface finishes of around 10- $\mu$ in. rms.

## REFERENCES

1. BISSON, E. E.; AND ANDERSON, W. J.: Advanced Bearing Technology. NASA SP-38, 1964.
2. McMULLAN, O. W.: High Temperature Materials and Their Heat Treatment for Anti-Friction Applications. SAE Preprint No. 122A, 1959.
3. MORRISON, T. W.; WALP, H. O.; AND REMORENKO, R. P.: Materials in Rolling Element Bearings for Normal and Elevated (450° F) Temperature. ASLE Trans., vol. 2, no. 1, 1959, pp. 129-146.
4. WACHENDORFER, C. J.; AND SIBLEY, L. B.: Bearing-Lubricant Endurance Characteristics at High Speeds and High Temperatures. Rept. No. AL65T068 (NASA CR-74097), SKF Industries, Inc., 1965.
5. BAMBERGER, E. N.: Bearing Fatigue Investigation. Rept. No. R67FPD309, General Electric Co. (NASA CR-72290), 1967.
6. ANDERSON, W. J.; AND ZARETSKY, E. V.: Rolling-Element Bearings. Machine Design-Bearings Reference Issue, vol. 40, no. 14, 1968.
7. ANDERSON, W. J.: Performance of 110-mm Bore M-1 Tool Steel Ball Bearings at High Speeds, Loads, and Temperatures. NACA TN 3892, 1957.
8. CARTER, T. L.: Preliminary Studies of Rolling-Contact Fatigue Life of High-Temperature Bearing Materials. NASA RM E57K12, 1958.
9. JACKSON, E. R.: Rolling Contact Fatigue Evaluations of Bearing Materials and Lubricants. ASLE Trans, vol. 2, no. 1, 1959, pp. 121-128.
10. WALP, H. O.; REMORENKO, R. P.; AND PORTER, J. V.: Endurance Tests of Rolling Contact Bearings of Conventional and High Temperature Steels Under Conditions Simulating Aircraft Gas Turbine Applications. WADC TR 58-392, 1959.
11. CARTER, T. L.; ZARETSKY, E. V.; AND ANDERSON, W. J.: Effect of Hardness and Other Mechanical Properties on Rolling-Contact Fatigue Life of Four High-Temperature Bearing Steels. NASA TN D-270, 1960.
12. BAMBERGER, E. N.; ZARETSKY, E. V.; AND ANDERSON, W. J.: Fatigue Life of 120-mm Bore Ball Bearings at 600° F with Fluorocarbon, Polyphenyl Ether, and Synthetic Paraffinic Base Lubricants. NASA TN D-4850, 1968.
13. ZARETSKY, E. V.; ANDERSON, W. J.; AND BAMBERGER, E. N.: Rolling-Element Bearing Life from 400° to 600° F. NASA TN D-5002, 1969.
14. CARTER, T. L.: A Study of Some Factors Affecting Rolling-Contact Fatigue Life. NASA TR R-60, 1960.
15. BAMBERGER, E. N.: The Effect of Ausforming on the Rolling Contact Fatigue Life of a Typical Bearing Steel. Trans. ASME, J. Lub. Tech., vol. 89, no. 1, 1967, pp. 63-75.
16. PARKER, R. J.; AND ZARETSKY, E. V.: Rolling-Element Fatigue Life of Ausformed M-50 Steel Balls. NASA TN D-4954, 1968.
17. ZARETSKY, E. V.; AND ANDERSON, W. J.: Rolling-Contact Fatigue Studies with Four Tool Steels and a Crystallized Glass Ceramic. Trans. ASME, J. Basic Engrg., vol. 83, no. 4, 1961, pp. 603-612.
18. ZARETSKY, E. V.; PARKER, R. J.; ANDERSON, W. J.; AND REICHARD, D. W.:



- Bearing Life and Failure Distribution as Affected by Actual Component Differential Hardness. NASA TN D-3101, 1965.
19. ZARETSKY, E. V.; PARKER, R. J.; AND ANDERSON, W. J.: Component Hardness Differences and Their Effect on Bearing Fatigue. *Trans. ASME, J. Lub. Tech.*, vol. 89, no. 1, 1967, pp. 47-62.
  20. ZARETSKY, E. V.; PARKER, R. J.; ANDERSON, W. J.; AND MILLER, S. T.: Effect of Component Differential Hardness on Residual Stress and Rolling-Contact Fatigue. NASA TN D-2664, 1965.
  21. SCOTT, R. L.; KEPPLER, R. K.; AND MILLER, M. H.: The Effect of Processing-Induced Near-Surface Residual Stress of Ball Bearing Fatigue. *Rolling Contact Phenomena*, J. B. Bidwell, ed., Elsevier Publ., 1962, pp. 301-316.
  22. GENTILE, A. J.; AND MARTIN, A. D.: The Effects of Prior Metallurgically Induced Compressive Residual Stress on the Metallurgical and Endurance Properties of Overload Tested Ball Bearings. ASME Paper 65-WA/CF-7, 1965.
  23. ALMEN, J. O.: Effects of Residual Stress on Rolling Bodies. *Rolling Contact Phenomena*, J. B. Bidwell, ed., Elsevier Publ., 1962, pp. 400-424.
  24. BUSH, J. J.; GRUBE, W. L.; AND ROBINSON, G. H.: Microstructural and Residual Stress Changes in Hardened Steel Due to Rolling Contact. *Rolling Contact Phenomena*, J. B. Bidwell, ed., Elsevier Publ., 1962, pp. 365-399.
  25. GENTILE, A. J.; JORDON, E. F.; AND MARTIN, A. D.: Phase Transformations in High-Carbon, High-Hardness Steels Under Contact Loads. *Trans. AIME*, vol. 233, no. 6, 1965, pp. 1085-1093.
  26. ZARETSKY, E. V.; PARKER, R. J.; AND ANDERSON, W. J.: A Study of Residual Stress Induced During Rolling. ASME Paper No. 68-Lub-1, 1968.
  27. BEAR, H. R.; AND BUTLER, R. H.: Preliminary Metallographic Studies of Ball Fatigue Under Rolling-Contact Conditions. NACA TN 3925, 1957.
  28. CARTER, T. L.; BUTLER, R. H.; BEAR, H. R.; AND ANDERSON, W. J.: Investigation of Factors Governing Fatigue Life with the Rolling-Contact Fatigue Spin Rig. ASLE Trans., vol. 1, no. 1, 1958, pp. 23-32.
  29. JONES, A. B.: Metallographic Observations of Ball Bearing Fatigue Phenomena. ASTM Symp. on Testing of Bearings, 1947, pp. 35-52.
  30. JOHNSON, R. F.; AND SEWELL, J. F.: The Bearing Properties of 1 Percent C-Cr Steel as Influenced by Steelmaking Practice. *Iron Steel Inst. J.*, vol. 196, pt. 4, 1960, pp. 414-444.
  31. KOVED, I.; AND ROSPOND, R. T.: Detection of Potential Fatigue Nuclei in Rolling Contact Bearings. SAE Paper No. 893B, 1964.
  32. ZARETSKY, E. V.; AND ANDERSON, W. J.: Evaluation of High-Temperature Bearing Cage Materials. NASA TN D-3821, 1967.
  33. PETERSON, M. B.; AND MURRAY, S. F.: Investigation of Possible Bearing Materials and Lubricants for Temperatures Above 1000° F. APEX 624, General Electric Co., April 1959.
  34. O'ROURKE, W. F.: Research on Developing Design Criteria for Anti-Friction Airframe Bearings for High Temperature Use. WADD TR 60-46, 1960.
  35. PARKER, R. J.; GRISAFFE, S. J.; AND ZARETSKY, E. V.: Surface Failure of Alumina Balls Due to Repeated Stresses Applied in Rolling Contact at Temperatures to 2000° F. NASA TN D-2274, 1964.
  36. PARKER, R. J.; GRISAFFE, S. J.; AND ZARETSKY, E. V.: Rolling-Contact Studies with Four Refractory Materials to 2000° F. ASLE Trans., vol. 8, no. 3, 1965, pp. 208-216.
  37. BAUGHMAN, R. A.; AND BAMBERGER, E. N.: Unlubricated High Temperature Bearing Studies. *Trans. ASME, J. Basic Engrg.*, vol. 85, no. 2, 1963, pp. 265-272.

38. RABINOWICZ, E.: Friction and Wear at Elevated Temperatures. WADC TR 59-603, 1960.
39. PARKER, R. J.; GRISAFFE, S. J.; AND ZARETSKY, E. V.: Surface Failure of Titanium Carbide Cermet and Silicon Carbide Balls in Rolling Contact at Temperatures to 2000° F. NASA TN D-2459, 1964.
40. WILSON, D. S.: Evaluation of Unconventional Lubricants at 1200° F in High-Speed-Rolling Contact Bearings. ASME Paper 61-Lub-9, 1961.
41. SIBLEY, L. B.; AND ALLEN, C. M.: Friction and Wear Behavior of Refractory Materials at High Sliding Velocities and Temperatures. ASME Paper 61-Lub-15, 1961.
42. TAYLOR, K. M.; SIBLEY, L. B.; AND LAWRENCE, J. C.: Development of a Ceramic Rolling-Contact Bearing for High Temperature Use. ASME Paper 61-Lub-12, 1962.
43. AMATEAU, M. F.; AND SPRETNAK, J. W.: Plastic Deformation of Magnesium Oxide Crystals Subjected to Rolling Contact Stresses. J. Appl. Phys., vol. 34, no. 8, 1963, pp. 2340-2345.
44. DUFRANE, K. F.; AND GLAESER, W. A.: Study of Rolling-Contact Phenomena in Magnesium Oxide. First Summary Rept., NASA CR-72295, 1967.
45. DUFRANE, K. F.; AND GLAESER, W. A.: Study of Rolling-Contact Phenomena in Magnesium Oxide. Second Summary Rept., NASA CR-72530, 1969.
46. ROBERTS, M.: A Study of Oil Circulation in the R4 Spin-Axis Bearing with Sintered Nylon Ball Retainer. Rept. E-2082, MIT Instrumentation Lab., December 1966.
47. ANON: Spin-Axis-Bearing Research. Rept. E-201, MIT Instrumentation Lab., January 1960.
48. ANON: Materials and Processes for High-g Gyro Application. AFAL-TR-65-91, MIT Instrumentation Lab., 1965.
49. ANON: Spin-Axis-Bearing Research. Rept. R-397, MIT Instrumentation Lab., 1963.
50. ANON: Vasco Matrix II CVM Alloy Steel. Alloy Digest, SA-200, August 1966.
51. LEMENT, B. S.; ALLEN, S.; AND KREDER, K.: Attainment of Higher Strength in Stainless Bearing Steels. Rept. E-2330, MIT Instrumentation Lab., September 1968.

# Effect of Materials—Metallurgy Viewpoint

**E. N. BAMBERGER**

**General Electric Company  
Cincinnati, Ohio**

This paper reviews the state-of-the-art of metallic materials technology as related to rolling-element bearings. In staying with the theme of the symposium, "Lubrication of Concentrated Contacts," the prime concern is therefore with those bearing structural materials which can operate within the capabilities of today's fluid lubricants.

The contributions of alloy development, alloy modifications, melting techniques, metalworking processes, and powder metallurgy procedures to improved bearing materials performance are discussed. It is shown that while satisfactory progress has been made in some of these technologies, other areas have lagged. This is due perhaps in part to a lack of stimulus by the customers but also due to the use of improper criteria for predicting the behavior of a material under rolling-contact conditions. Lastly some potential areas for future research in bearing materials technology are discussed.

**T**HE ADVENT of the aircraft gas turbine markedly and dramatically changed existing concepts regarding transportation. No less dramatic has been its effect on the metallurgist's concept of structural materials for aircraft applications. The continuously increasing thermal and mechanical demands the jet engine has made on materials has challenged the metallurgical community more than any other single event in its history.

The unique aspect of this materials revolution is the broad scope of materials technology affected. High-temperature superalloys, titanium, composites, plastics, beryllium, are only a few examples of the many areas of materials technology which were either created in response to the demands of the jet engine or whose development was accelerated because of it. Similarly the processing of metallic materials has been critically reexamined and in many instances drastically altered. Specialized melting, forging, and joining processes, for example, have been developed to meet jet engine requirements.

Included in this materials revolution has been that segment dealing with rolling-element bearing metallurgy, lubrication, design, and analyses.



It is safe to say that modern bearing metallurgy and lubricant development had its start with the dawn of the jet age, and the growth of these two disciplines parallels that of the aircraft gas turbine.

The purpose of this chapter is to review briefly the present state-of-the-art of rolling-element metallurgy and to indicate some of the areas that will need to be explored in the future. It is relatively simple to trace the growth of modern bearing metallurgy since it is very recent history. For example, in one of the classic textbooks on bearing analysis (ref. 1) published in 1949, the only rolling-element bearing material discussed is SAE 52100. Even as recently as 1957, another authoritative source (ref. 2) makes only incidental note of the fact that SAE 52100 is not useful over 350° F and that "For temperatures above 350° F, bearing manufacturers have made small lots of bearings of M-1 and M-10 tool steels. These steels retain their hardness to temperatures approaching 1000° F. Evidence available to date indicates that they operate satisfactorily provided lubrication can be maintained." Perhaps this is stretching a point, because in 1957 materials such as M-1 and M-50 were finding their way into critical jet engine applications. Nevertheless, it does emphasize the relatively brief history of tool steels in bearing applications.

SAE 52100 has been the prime bearing material since about 1920, and even today constitutes probably 90 percent of bearing production. However, as engine operating temperatures increased, it was recognized that materials with better hot hardness and higher tempering temperatures were required. In the search for materials with these properties, alloy steels initially developed for tool applications were investigated. Some of these materials investigated are shown in table 1. All of these materials are through-hardenable, which means hardness is achieved by cold working or by heat treatment rather than by a surface modification such as carburizing or nitriding.

Fortunately several of the materials such as M-1 and M-50 were found to be satisfactory for rolling-element bearing use, and M-50 is today the material used by most American jet engine manufacturers. The word "fortunately" is used because, with the exception of WB-49, none of the materials were developed specifically for bearing applications.

In figure 1, which shows the elevated temperature hardness of some bearing steels, it may be seen that several materials exhibit better high-temperature characteristics than M-50. The question naturally arises, why not use these materials? The answer to this is two-fold. First, a vast amount of experimental and operational data has been generated on M-50. Consequently the confidence factor on this material is very high, and jet engine bearing metallurgists are generally satisfied with its high-temperature performance. The second and perhaps more important reason is the engine designers' development of sophisticated cooling procedures that have kept the bearing and oil temperatures at levels far below

TABLE 1.—*Nominal Composition of Materials for Rolling-Element Bearings*

Bearing steel designation	Maximum operating temperature <sup>a</sup> (°F)	C	P (max.)	S (max.)	Mn	Si	Cr	V	W	Mo	Co
SAE 52100	350	1.00	0.025	0.025	0.35	0.30	1.45	...	...	...	...
MHT <sup>b</sup>	500	1.03	0.025	0.025	0.025	0.35	1.50	...	...	...	...
AISI M-50	600	0.80	0.030	0.030	0.30	0.15	4.00	1.00	...	4.25	...
AISI M-10	800	0.85	0.030	0.030	0.25	0.30	4.00	2.00	...	8.00	...
AISI M-1	900	0.80	0.030	0.030	0.30	0.30	4.00	1.00	1.50	8.00	...
AISI M-2	900	0.83	0.030	0.030	0.30	0.30	3.85	1.90	6.15	5.00	...
WB-49	1000	1.07	0.006	0.007	0.30	0.02	4.40	2.00	6.80	3.90	5.2
52Cb	450	0.85	...	...	0.35	0.85	0.90	...	...	0.60	...

<sup>a</sup> Based on hot hardness. Susceptibility to certain corrosive and oxidative environments may reduce the effective maximum temperature level.

<sup>b</sup> Also contains 1.36% Al

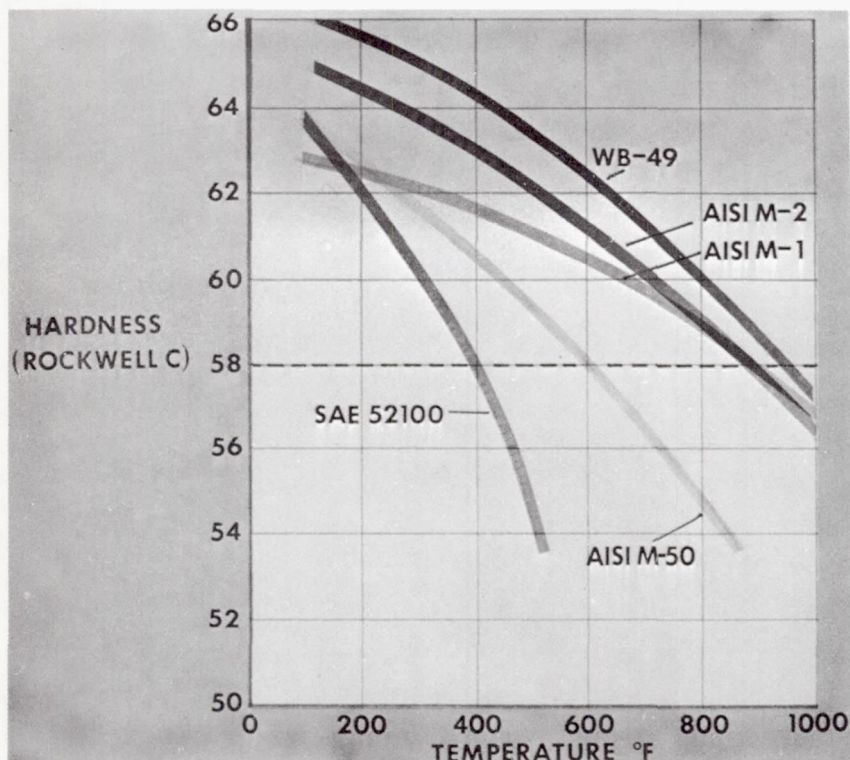


FIGURE 1.—Hardness of various bearing steels as a function of temperature.

those previously anticipated. In a conference held in 1960 (ref. 3), predictions of oil (and bearing) temperatures reaching 500° F in 1968 were made. Yet today, 9 years later, bearing temperatures of most current and advanced aircraft are not much higher than engines of 1960 vintage, despite the fact that thrust levels have increased by as much as a factor of 4.

The advances in engine cooling techniques have greatly influenced the direction of the current effort in bearing metallurgy. The major concern today is with increased bearing life and reliability rather than increasing the temperature capability. The jet engine of today and those planned for the near future will operate for longer total times, longer times between overhaul, and with a higher degree of reliability than ever before. The bearings are a vital factor in achieving these objectives.

Rolling-element bearing failures can generally be attributed to the following factors:

- (1) Rolling-contact fatigue—10 percent



- (2) Lubricant problems—85 percent
- (3) Others—5 percent

Of these, the only failure mode over which the metallurgist can exercise any control is fatigue. In this task he can either change material or improve the existing material. Changing materials involves developing new materials or attempting to apply already existing ones, the latter approach having been used in introducing the family of molybdenum-base tool steels currently in use.

The second alternative is to modify or improve existing materials. This is where the major effort has been and is being concentrated in current bearing materials technology. Most such modifications consist of either improving the metallurgical processing used to make a given alloy, modifying the alloy composition, or changing the processing of the material. The major areas of interest therefore are alloy development or modification, melting practices, processing techniques, and powder metallurgy.

Before reviewing some of the details of the above, it seems appropriate to touch briefly upon one item which to some extent has had a negative effect on all of these developmental activities, particularly the alloy development portion. Rolling-contact fatigue is a unique property and consequently is not directly relatable to such standard material mechanical properties as tensile, creep, and rupture strengths. One exception may be hardness which, as shown in figure 2, is usually a reasonably reliable indicator of bearing fatigue, although even this is not always the case. For example, in the case of certain unique processing techniques such as ausforming, significant improvements in fatigue life are achieved without an increase in hardness.

In the past, much of the bearing alloy improvement effort was geared to achieving improved standard mechanical properties such as the ones noted above. When materials thus developed were applied to and tested in actual bearings, the results were often disappointing since bearing performance simply does not correlate well with the traditionally measured properties.

Only in recent years has a good definition of those mechanical and physical properties significant to rolling-element bearing performance been established. These are listed below:

- (1) Rolling-contact fatigue life
- (2) Hardness—room and elevated temperatures
- (3) Coefficient of expansion
- (4) Fracture toughness
- (5) Corrosion resistance
- (6) Metallurgical transformation characteristics

By evaluating candidate materials on the basis of these pertinent prop-

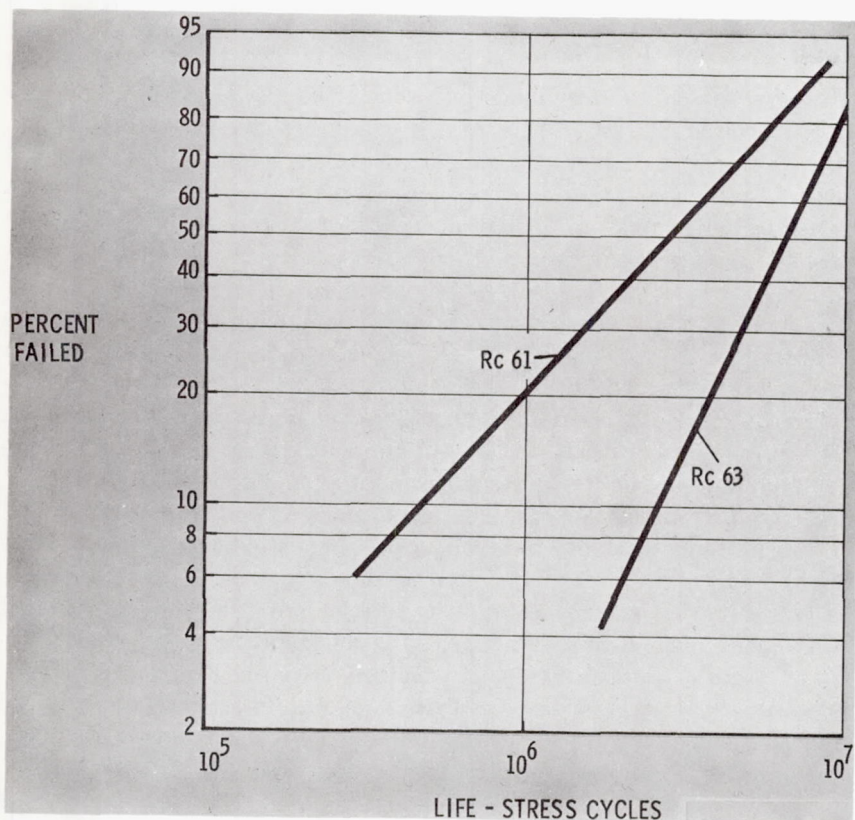


FIGURE 2.—Effect of hardness on rolling-contact fatigue life.

erties, the metallurgist should be able to obtain a valid judgement as to the ultimate efficacy of the material in an actual bearing application.

Obviously a standard should be available against which to compare the new or modified materials. Since consumable vacuum-melted M-50, on the basis of millions of jet engine flying hours, has proved to be an excellent bearing material, it qualifies as the logical choice for such a standard. Naturally, even with this type of a yardstick, there are trade-off situations that will arise and need to be individually considered. A case in point might be a new alloy that has somewhat lower rolling-contact fatigue life than M-50 but evidences superior corrosion resistance under certain environmental conditions. For a specific application, this alloy might well qualify as the first choice of the designer. In any event, unless realistic criteria are used to evaluate candidate materials for bearing applications, progress in developing new steels or improving existing ones will continue to be a slow and uncertain process.



### ALLOY DEVELOPMENT AND MODIFICATION

With very few exceptions, alloy development activities have not contributed significantly to the availability or improvement of rolling-element bearing steels. Of those materials listed in table 1, only WB-49 was specifically developed for bearings (ref. 4). Its use has been limited for economic reasons and because its performance is marginal when evaluated in large-diameter bearing sizes (ref. 5), but mainly because M-50 has proved to be an eminently satisfactory high-temperature material. In a series of high-temperature bearing tests, M-50 was operated at 600° F with no reduction in bearing dynamic capacity (ref. 6).

Another material, now identified as WDC-65, was also developed for bearing applications (ref. 7), although its use has been hindered by forging difficulties due to its high alloy content. There is promise, however, that powder metallurgy consolidation techniques can be applied to this material, thus making it available for evaluation and potential application. A number of other materials have been cited as candidates for bearing use (refs. 8 to 15), although none of these found widespread acceptance.

The bearing alloy development work performed in the past has suffered to some extent because of the previously mentioned problem of not using the proper criteria for judging the performance of a material for bearing applications. Work currently being performed in this area is more correctly aimed at meeting the pertinent bearing material properties, and consequently some new alloys can be expected in the next few years which should offer sizeable improvements over those available today.

Alloy modification efforts have been somewhat more fruitful than alloy development activities. This area also has been more productive in advancing the general field of bearing material physical metallurgy. Studies have been performed relating the effect of various modifications in chemistry on bearing steel properties (refs. 16 to 20). Even here, only a few of the investigators have evaluated the modifications in terms of the specific requirements for rolling-contact fatigue. The alloy modification field has, however, produced 52Cb (ref. 21), a modified SAE 52100 (primarily a higher Mo content) which has found some application in today's jet engines. Another SAE 52100 derivative, MHT (52100 + 1.36% Al) has also found limited use in certain engine bearings.

### MELTING PRACTICES

Compared to the alloy development work, considerable improvements in bearing steels have been made by the application of new melting techniques. In the earlier stages, much attention was given to the effect of acid vs basic refractory air-melting methods (refs. 22 to 24). Even



today data are still being presented (ref. 25) that show the superiority of acid-refractory produced material for bearing applications.

The major advances in melting practice in the past 10 to 15 years have come with the introduction of vacuum-melting procedures. In fact, the use of vacuum melting has markedly influenced the quality and strength levels of most of the high-strength structural materials used in the jet engine industry, as illustrated in figure 3.

Vacuum processing reduces or eliminates the amount of non-metallic inclusions, entrapped gases, and trace elements present in structural alloys. This results in substantially cleaner material. The two prime methods of vacuum processing are vacuum induction melting (VIM) and consumable-electrode vacuum melting (CVM).

In vacuum induction melting, a cold charge is melted in a high purity refractory crucible in an induction furnace and subsequently poured into statically cast ingots, the whole operation being performed while the melt is under a vacuum. This is schematically illustrated in figure 4.

Consumable-electrode vacuum melting (CVM) is a technique wherein electrodes made from a primary air or vacuum induction melted heat are remelted by an electric-arc process. The remelted product solidifies in a water-cooled copper mold under vacuum as illustrated in figure 5.

The CVM method produces cleaner and more segregation-free material than the VIM process. This is partially due to the fact that a refractory crucible is used in the induction-melting process with the result that refractory particles or intermetallic compounds formed by metal-refractory reactions occasionally get into the ingot. In the CVM process the ingot is used as one terminal of the arc and is melted into a water-cooled copper crucible. The solidification pattern of the CVM

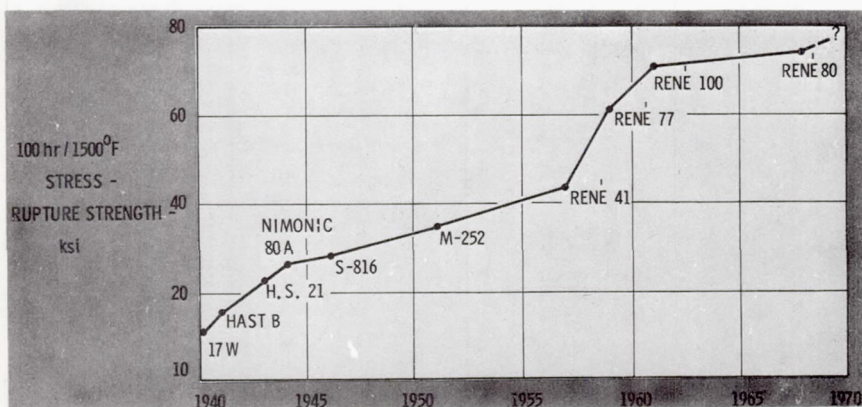


FIGURE 3.—Evolutionary increases in high-temperature alloy strength through alloy development and vacuum melting (circa 1957) techniques.

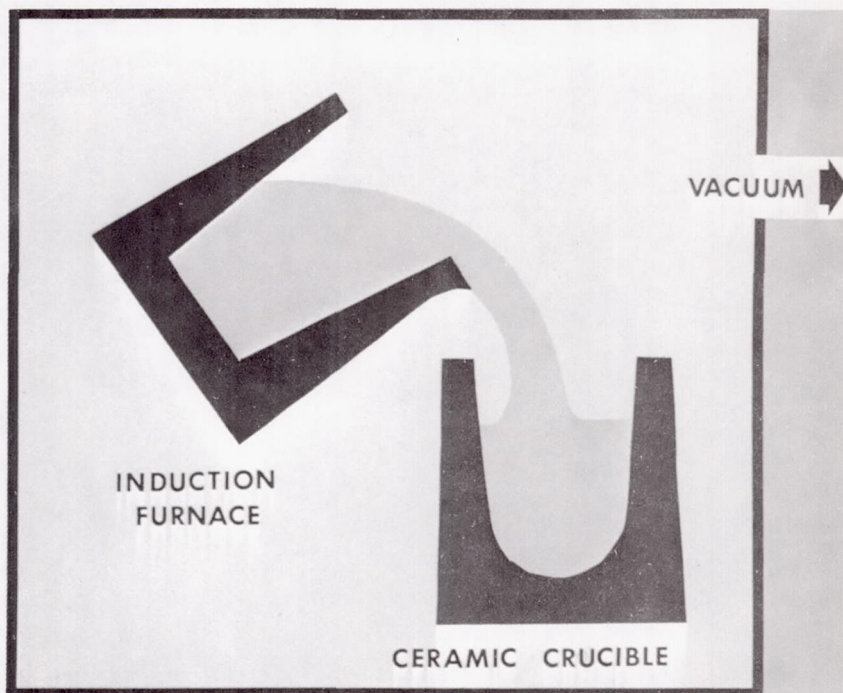


FIGURE 4.—Schematic of induction vacuum melting process.

ingot is also totally different from that produced by VIM, since only a shallow molten pool is present during CVM and little of the "pipe" or segregation characteristic of statically cast ingots occurs. This has a profoundly beneficial effect on the homogeneity and consistency of the steel. While VIM and CVM are the prime methods used today to produce materials such as M-50, other vacuum-process methods have been developed, primarily aimed at improving 52100 (ref. 26).

Electro-slag melting or remelting (ESR) has been used extensively in Europe and Russia (refs. 27 to 34) for improving bearing steels. The ESR process is a consumable-electrode remelting procedure in which a molten flux blanket is maintained over the molten metal. The flux blanket serves essentially the same purpose as a vacuum atmosphere in protecting the ingot from contamination.

In this country, the ESR process has been proven to yield results similar to that achieved by CVM for many iron and nickel alloys. This is logical since solidification again occurs in a water-cooled copper mold containing a shallow molten pool. Heat input occurs by conduction from the resistance-heated slag rather than by arcing of the remelt

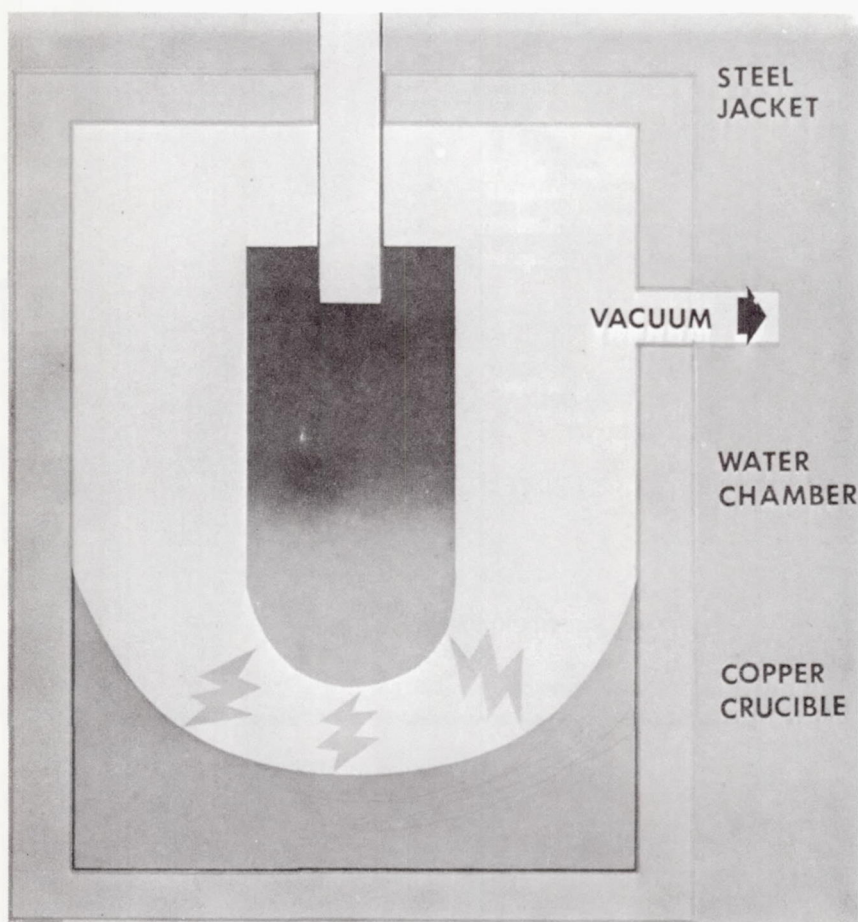


FIGURE 5.—Schematic of consumable-arc vacuum remelt process.

electrode. Early variations of the process in this country were referred to as the "Hopkins" or "Kellogg" process. To date some M-50 has been ESR-processed, and indications are that bearing life is equivalent to CVM M-50. It also appears that ESR-processed material may be more workable due to a reduced incidence of macro-segregation.

Electron-beam (EB) melting has also been used to produce ultra-high-purity bearing steel (ref. 35). The bulk of this has been done in the Soviet Union and East Germany where the equipment is more common than in the United States. Since EB melting is performed under a very hard vacuum, it produces alloys similar in cleanliness to those made by VIM or CVM. Solidification also occurs in a shallow



pool similar to CVM or ESR. Until recently, facilities for EB melting in this country were restricted to relatively small billet sizes. Larger-capacity equipment, however, is becoming available and should make this process economically more attractive.

It is possible with any of these melting techniques to produce material with a lower inclusion content than air-melted material, particularly those inclusions that are considered to be most injurious such as oxides, silicates, and aluminates. These inclusions are, in part, the result of standard air-melt deoxidation practice which involves the use of silicon and aluminum. Exposing the melt to a vacuum permits deoxidation to be performed effectively by the carbon. The products formed when using carbon as a deoxidizer are gaseous, and thus escape within the vacuum leaving no harmful residue in the melt. Furthermore, these techniques permit extremely close control of chemistry and also permit production of variations in chemical analysis which were at one time impractical.

Even in exceptionally clean materials, nonmetallics are present to some degree. Depending on their magnitude and location in relation to the contact stresses, they can act as the nuclei for fatigue cracks. There are, however, still no definitive data to indicate the exact type of inclusion that is detrimental to rolling-contact fatigue life. Obviously the hard brittle inclusions such as oxides and silicates are suspect, although even here the shape of the inclusion may have a major influence. The effect of inclusions has been studied by a number of investigators (refs. 36 and 37). While they show that vacuum melting undoubtedly improves fatigue life, this improvement is not always commensurate to the improvements made in the cleanliness of the steel. On the other hand, it has been shown that repeated CVM remelting markedly improves fatigue life (ref. 20). Bearing endurance results, summarized in figure 6, show that the  $B_{10}$  life (life at which 90 percent of a group of bearings will survive) appears to increase for successive remelting with the fifth remelt material reaching a life of four times the original air-melt group.

A totally different approach to the problem of nonmetallic inclusions are the studies being performed to determine the effect of increasing the sulfur content of bearing steels (ref. 38). Sulfur, as such, is an inclusion former; however, it has been shown that sulfurizing by lowering the oxygen content lowers the amount of hard and brittle oxide type inclusions in the steel. Consequently the improvement in rolling-contact fatigue tests performed on sulfurized SAE 52100 steel are thought to be due to the reduction of potential stress-concentration sites, thereby reducing the probability of crack initiation. A further reduction of effective stress raisers is also caused by the encapsulation of the oxide-type inclusions by sulfides.

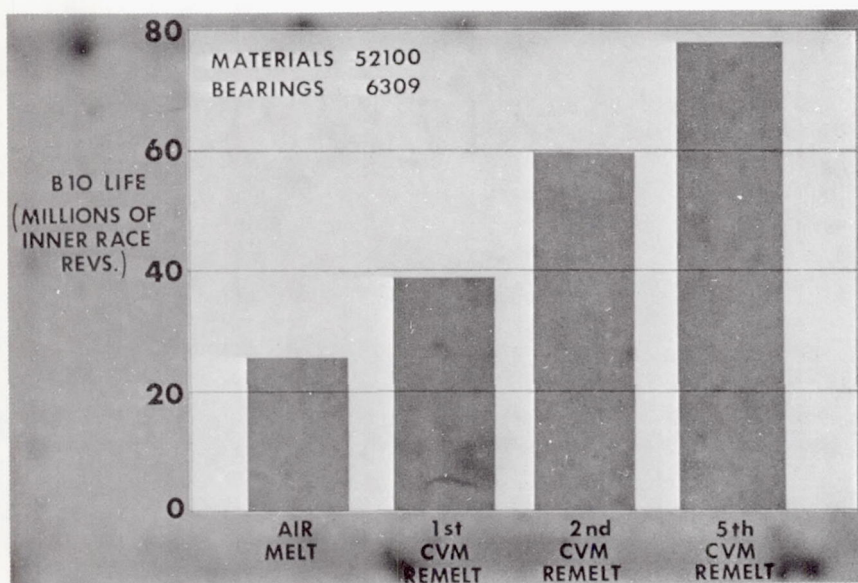


FIGURE 6.—Effect of multiple consumable-arc vacuum remelting on bearing fatigue life.

The amount of sulfur being added is at present relatively small, i.e., generally only as high as the upper limit (0.025 percent) of most bearing material specifications. Some heats, however, have been made with a sulfur content of 0.043 percent and some as high as 0.1 percent. If this modification should prove beneficial, another advantage could be an improvement in the machinability of bearing steels as the addition of the sulfur promotes a more free-machining steel.

#### PROCESSING TECHNIQUES

The third major area of interest, processing techniques, has also aided significantly in achieving improved bearing materials. Considerable work has been performed on the effect of residual stresses on fatigue life and the mechanisms for introducing these into the material.

Stresses can be induced in a material by producing microscopic and macroscopic deformations that put residual stresses on a unit volume of material. Residual stress can either increase or decrease the maximum shearing stress:

$$(\tau_{\max})_r = \tau_{\max} - \frac{1}{2}(\pm S_r)$$

where  $\tau_{\max}$  is the maximum shearing stress,  $(\tau_{\max})_r$  is the maximum shearing stress modified by the residual stress, and  $S_r$  is the residual stress, the positive or negative sign indicating a tensile or compressive residual stress, respectively. A compressive residual stress reduces the



maximum shearing stress and increases fatigue life according to the inverse relation of life and stress.

Residual stresses can be induced by heat treating, rolling, shot peening, diamond burnishing, and severe grinding. Each of these methods (except heat treating) is a separate mechanical process performed subsequent to heat treating.

The General Motors Research Laboratories have performed work which indicates that these compressive residual stresses do indeed increase fatigue life (ref. 39). Based on a test program of 50 bearings, the group having compressive residual stresses (greater than 140 000 psi) exhibited approximately a 2 to 1 increase in life over those with no residual stresses. In other tests, life differences on the order of 3 to 1 were achieved.

By diffusing nitrogen gas into the surface of a ball or race, General Motors Research Laboratories were able to lower the temperature at which austenite transforms to martensite in the surface areas. Then, by avoiding large temperature gradients during quenching, the interior of the part can be made to transform to martensite before the surface zone reaches its martensitic transformation temperature.

As a result of this manipulation, the surface zone, which remains in the austenitic state and thus quite soft and ductile, readily accommodates the volume expansion of the interior as it transforms from austenite to martensite. Since the surface yields by plastic deformation, no residual stresses result in the interior. (Normally, residual compressive stresses result when expansion occurs against the hardened exterior.) When the surface zone subsequently transforms into the martensitic state, the accompanying expansion is not accommodated by the already hardened interior, with the result that there is no plastic deformation of the interior regions. Thus the expansion of the surface during the austenite-martensite transformation results in residual compressive stresses on and just below the surface of the ball or race, and ideally in that region where it will have the greatest benefit on fatigue.

Another processing technique which has been used to improve bearing life is the manipulation of the material fiber orientation. The races and rolling elements of most bearings are forged. Any metallic object formed by forging generally possesses a fiber-flow pattern which reflects the flow of metal during the forging operation. Carbides and nonmetallic inclusions are progressively and directionally oriented during each forming operation from the ingot to the final bearing-element shape. Although the desired microstructure is obtained by heat treatment, the carbides and inclusions generally retain their processing-induced directionality. This pattern is fibrous in appearance when the part is macro-etched, hence the term "fiber-flow lines." In addition, the entire grain pattern is preferentially oriented in the same manner.



The type of forging used to produce rolling-element bearing components will determine the fiber pattern. Steel balls are usually fabricated by upsetting between hemispherical dies. This technique produces a fiber-flow pattern with two diametrically opposed areas with fibers oriented approximately perpendicular to the surface, as illustrated in figure 7. These areas are commonly known as the poles. NASA has evaluated the influence of these end-grained polar regions on rolling-contact fatigue (ref. 36). Their data show that a very significant increase in failure density occurs in the polar areas, and a small increase in failure density also occurs at the equator where a thin band of perpendicular fiber exists.

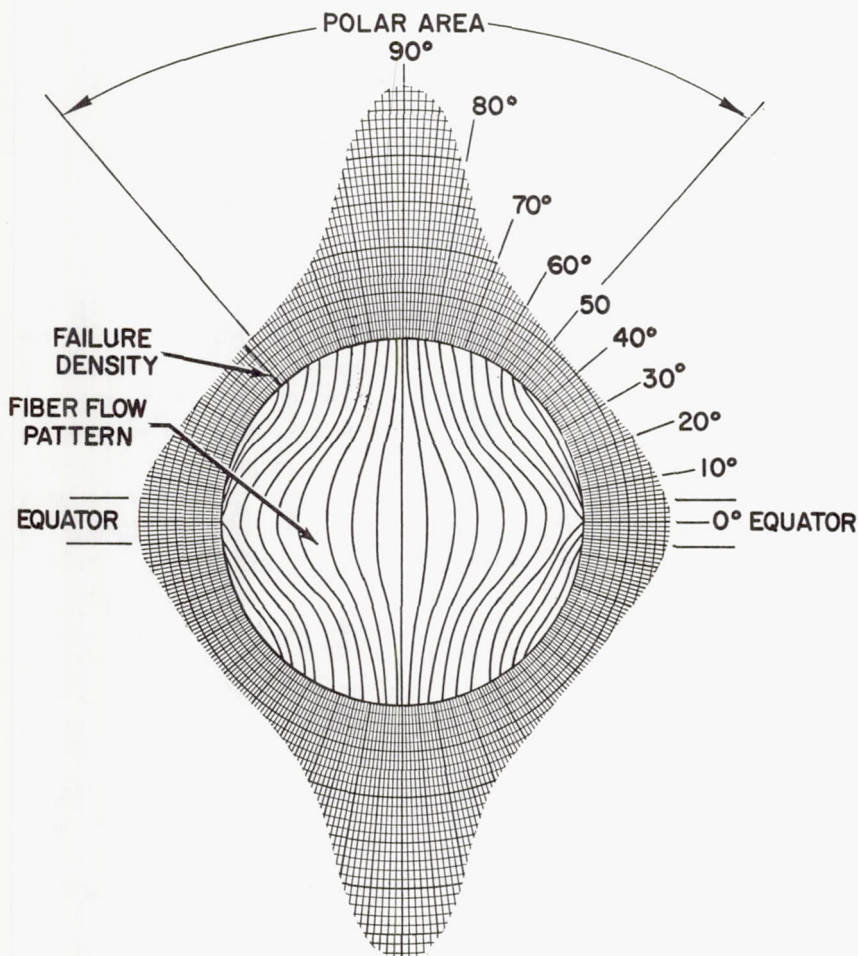


FIGURE 7.—Failure density of an upset ball as a function of fiber orientation.

Although the economic feasibility of controlling fiber flow in balls is questionable, this is not the case with bearing races. The Fafnir Bearing Company has devised a forging method to induce parallel fiber orientation in bearing races as shown in figure 8. The method consists of upsetting the end of a bar to form a bulge, then punching out the bar, leaving a ring. It is customary to contain the upset steel in dies that can be shaped to help flow the steel in the direction desired. A bearing race made by a more conventional method is shown in figure 9. In this case the steel was upset to a pancake, the center was punched out, and the ring was then expanded by hammering or rolling. A ring of rectangular cross-section is formed by this method with a curving, bulging grain flow nearly perpendicular to the race way.

Bearings having side grain or fiber flow parallel to the race have shown a significant increase in life over bearings with the end grain or fiber nearly perpendicular to the running track. Also, life of parts seems less sensitive to steel quality variations when forged with the fiber flow parallel to the race.

Within recent years a thermo-mechanical process termed "ausforming" has been evaluated which has shown great potential in improving bearing fatigue life. The term "ausforming," "ausformed," and the like, are used to describe processing techniques which consist of the working of a steel while the material is in the metastable austenitic condition.

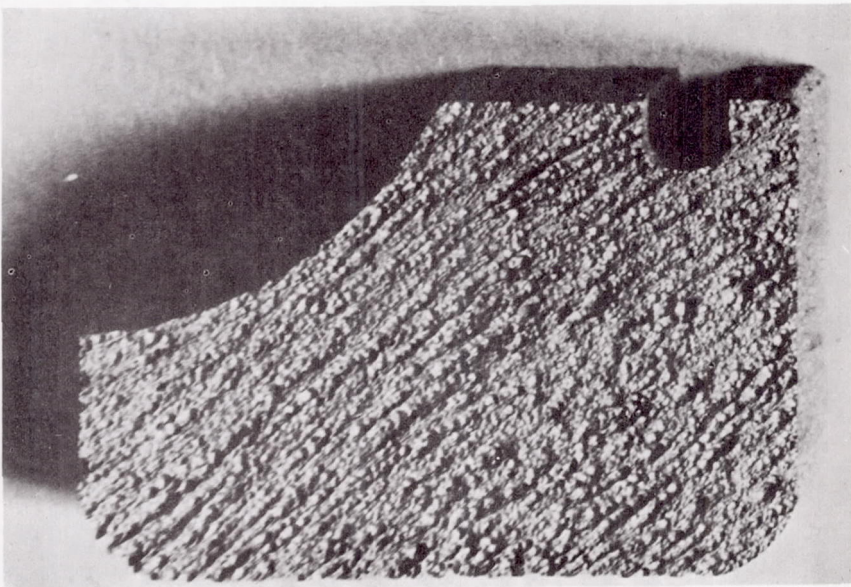


FIGURE 8.—Forged bearing race showing fiber flow parallel to raceway.



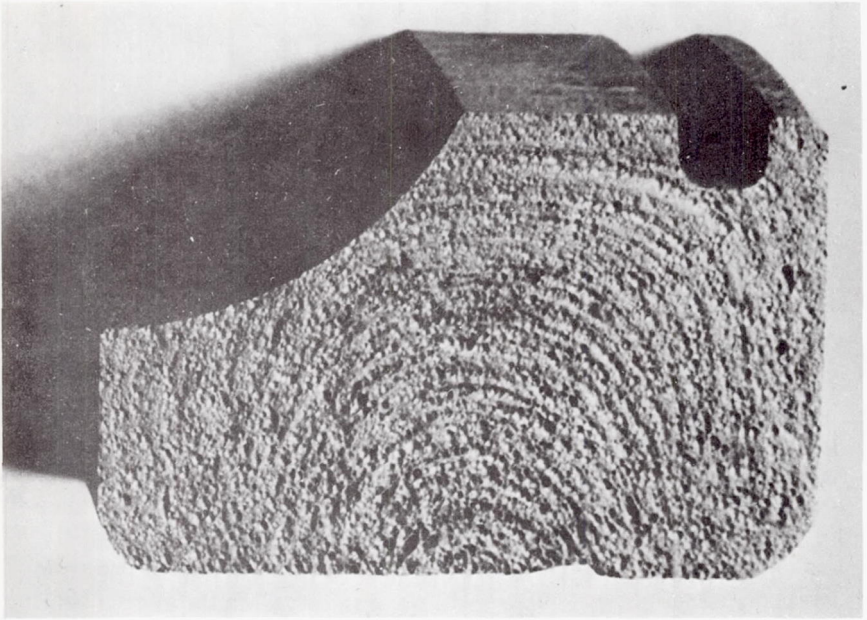


FIGURE 9.—Bearing race showing fiber flow primarily perpendicular to raceway.

Ausforming has been studied since 1954 when Lips and Van Zuilen first reported on work which they had been performing (ref. 40). Since then a number of organizations both in the U.S. (refs. 41 to 43) and abroad (ref. 44) have investigated the process. The application of ausforming to rolling-element bearings was first reported in 1965 (ref. 45). The material used in this work was M-50 because this alloy has the required metallurgical transformation characteristics for ausforming, as illustrated in figure 10. The upper curve is the transformation diagram for AISI 52100. In this material the phase changes occur rapidly upon quenching from the austenite range, and there is not sufficient time to effect any type of deformation in the austenite prior to its transformation to martensite. The bottom curve illustrates the transformation behavior of an AISI M type tool steel where the austenite is isothermally stable for a sufficient length of time to work the material.

In the initial studies, M-50 bar stock was ausformed to 40-, 70-, and 80-percent deformation. To obtain a valid comparison to standard heat-treated material, it was necessary to assure that the hardness of the ausformed material was identical to the standard M-50. This required a tempering study, during which it was noted that the 80-percent deformed material did not exhibit the secondary hardening behavior common to this type of steel. This is illustrated in figure 11. The absence



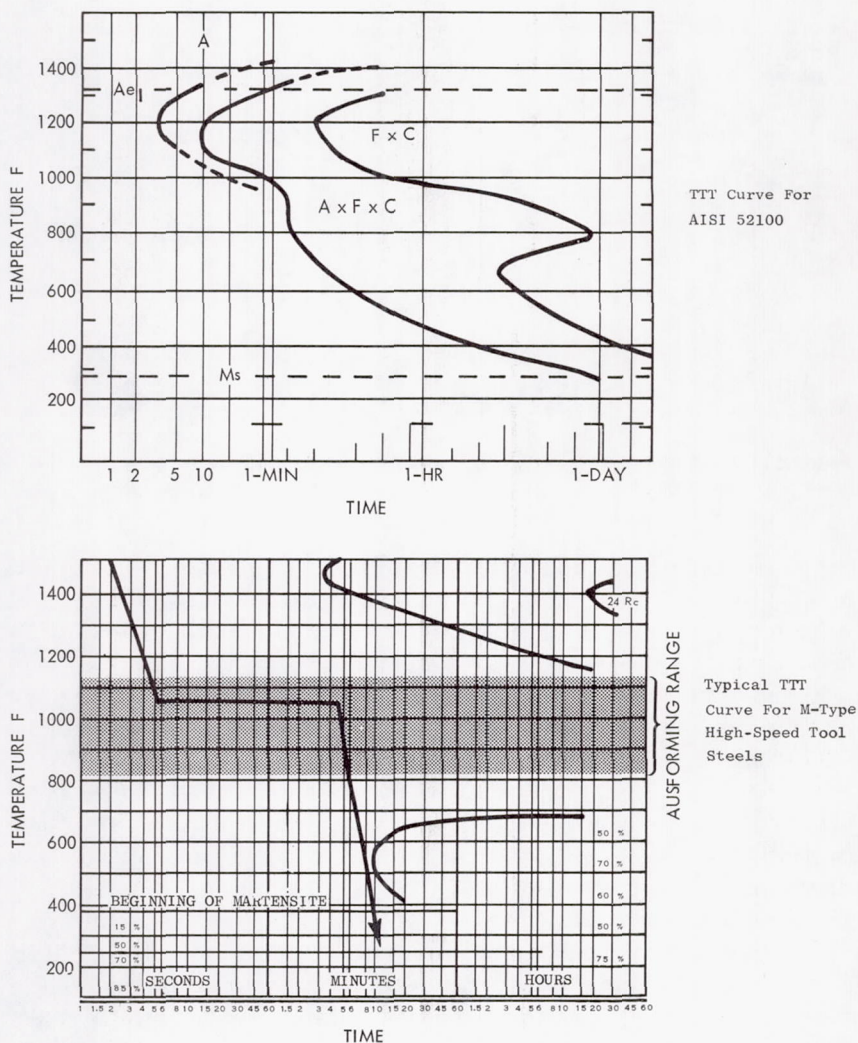


FIGURE 10.—Time-temperature transformation diagrams for two typical bearing steels.

of secondary hardening was observed only in the material having the 80-percent reduction, whereas with lesser amount of deformation, some evidence of secondary hardening was observed. This is significant as it indicates a relatively simple means of determining whether the required amount of work has been induced into the material.

Testing was performed on a rolling-contact fatigue rig similar to the one shown in figure 12. The family of curves obtained on the ausformed material are shown in Weibull distribution in figure 13. The curve for

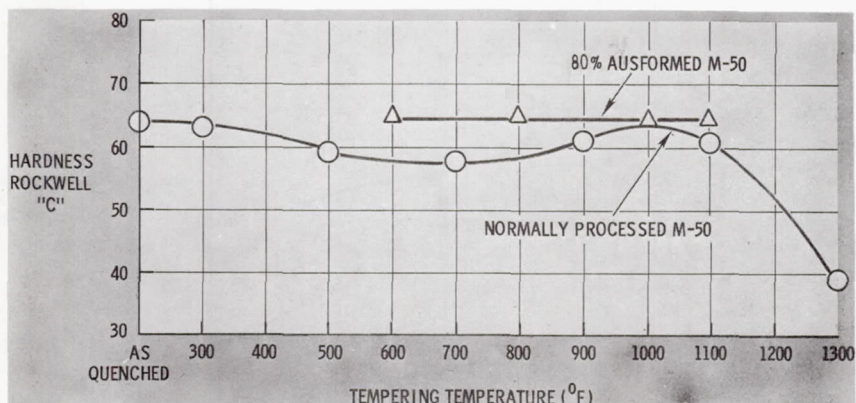


FIGURE 11.—Tempering response of ausformed M-50 to standard heat-treated M-50, indicating absence of secondary hardening in ausformed material.

the 80-percent worked material, which showed the greatest improvement in life, is reproduced in figure 14. This is compared to the results obtained when testing bars are made from the same heat of material but normally heat treated. The 90-percent confidence band for this latter series of tests indicates that the ausformed material constitutes a separate and superior population from the standard heat treated material with an improvement of over 600 percent in the  $B_{10}$  life.

The next step was to evaluate the effect of ausforming in actual bearing tests. The bearing selected for this study was a 35 mm bore single row radial ball bearing. Twenty such bearings having ausformed inner rings and balls and standard heat treated outer rings were tested (ref. 46). The test results of the ausformed bearings are presented in figure 15. For reference, the curve on the left was obtained by testing 27 bearings having all standard heat-treated CVM M-50 components. These bearings were identical to the ausformed bearings and were tested under the same conditions. The results confirmed the improvement in rolling-contact fatigue life predicted by the rolling-contact rig tests.

There are a number of theories proposed for the strengthening mechanism in ausforming. These are

- (1) Strain-induced precipitation
- (2) Reduction in martensite platelet size
- (3) Directional properties—grain flow
- (4) Cold-work strengthening of austenite
- (5) Breakup of large nonmetallic inclusions

Since a very specific mechanical property, i.e., rolling-contact fatigue, is being affected, it appears that strain-induced precipitation and its

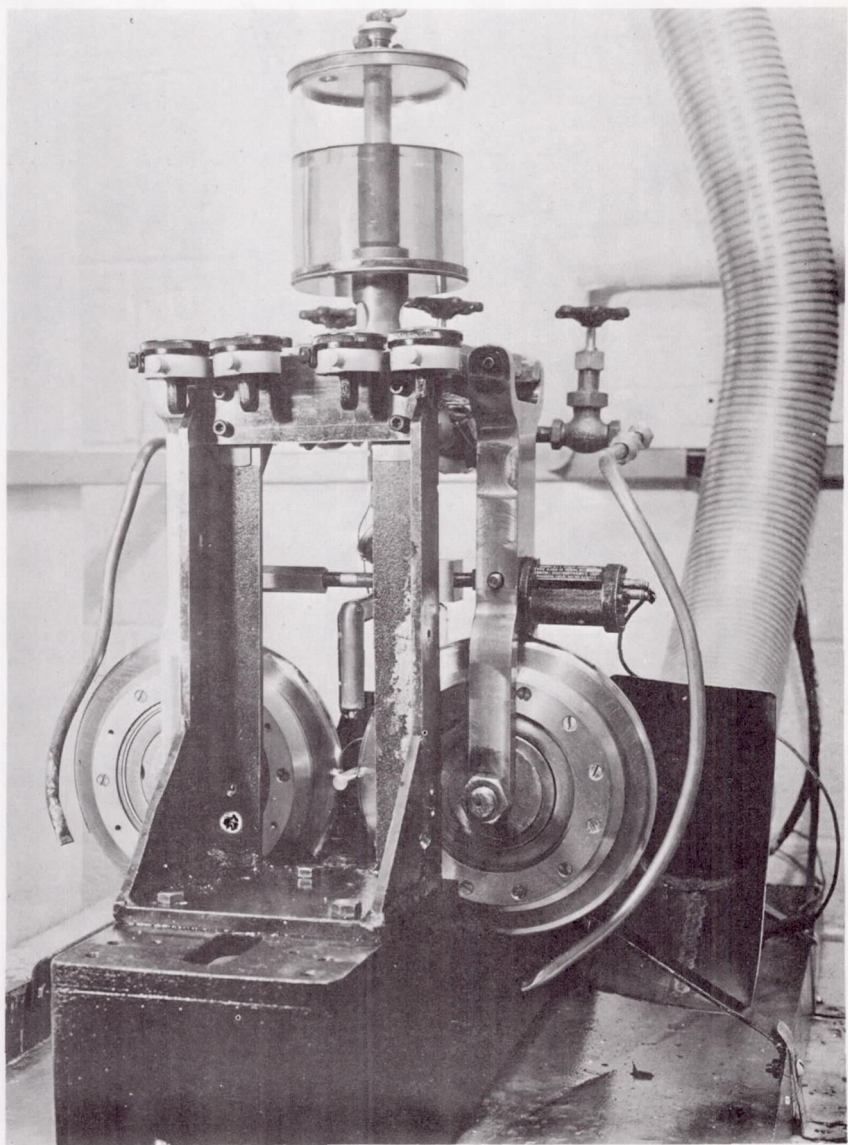


FIGURE 12.—Rolling-contact fatigue tester.

interplay with strain hardening is the predominant beneficial factor in this case.

The transformation kinetics for the standard heat-treated material are well established and briefly are as follows. At the austenitizing temperature ( $2050^{\circ}$  F for M-50), much of the carbon- and carbide-



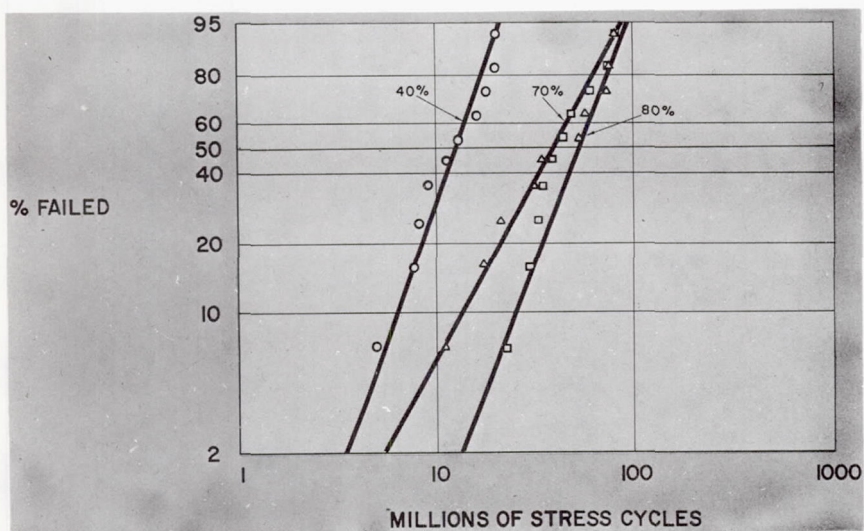


FIGURE 13.—Effect of percentage deformation on rolling-contact fatigue life of ausformed M-50.

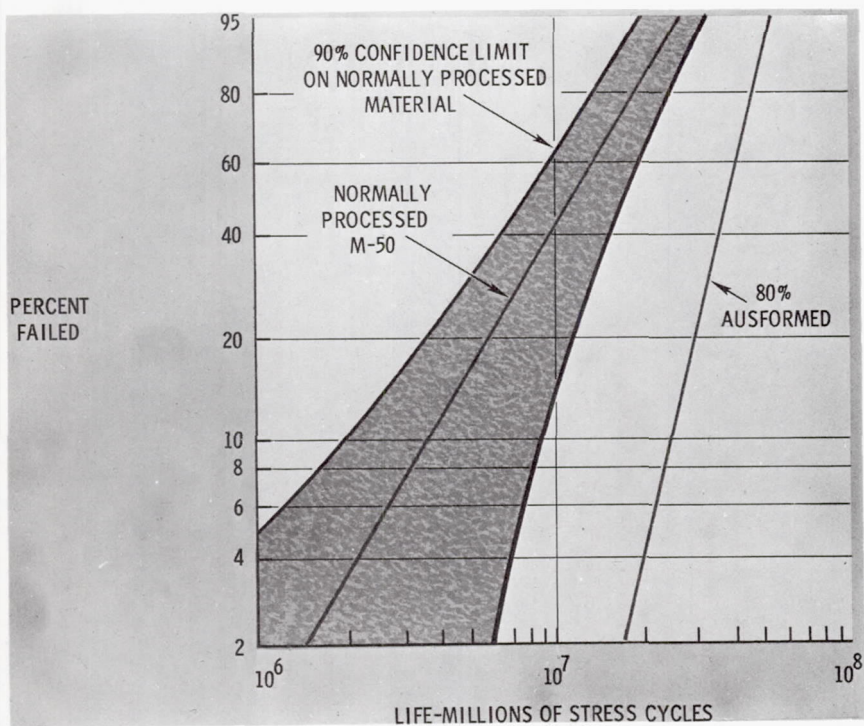


FIGURE 14.—Results of rolling-contact fatigue tests showing the effect of ausforming on the fatigue life of M-50.

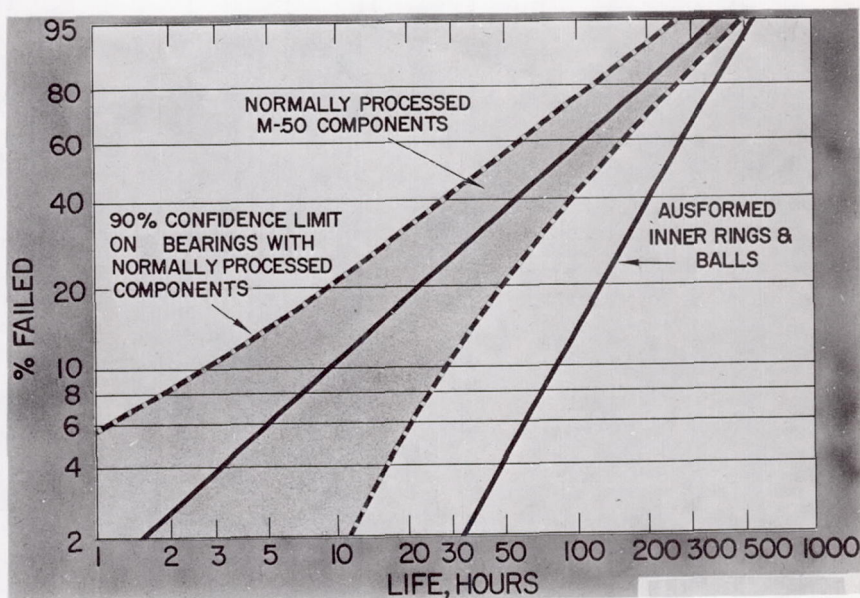


FIGURE 15.—The effect of ausforming on bearing life.

forming elements are rendered in solid solution. A rapid quench to the martensite range results in a super-saturated solution which upon tempering results in the precipitation of carbides in the unstrained matrix.

In the ausforming process, however, sufficient energy (strain) and time is given to the material so that carbide precipitation can occur at the ausforming temperature. Since carbide nucleation sites and diffusion rate are expected to be greatly increased, many more carbides are formed with the result that they are smaller and more uniformly dispersed. Also, as a result of this increased dispersion, much more of the carbide-forming alloying elements are utilized with the end result that the amount remaining in solution is apparently insufficient either to give rise to secondary hardening or to impart the normal etching characteristics to the martensitic matrix.

The amount of deformation is an influential factor, as it dictates at what step in the mechanical and thermal processing the precipitation of alloy carbides occurs. In the heavily (80 percent) deformed material, the absence of secondary hardening indicates that carbide formation occurs at the temperature and during the time which the material is being worked. In those materials having lesser amounts of deformation, it is thought that most of the carbide precipitation still occurs during the tempering cycle as evidenced by secondary hardening during tempering. This is consistent with the strain-induced precipitation theory,



as the role of this mechanism is to provide the numerous nucleation sites during the deformation process, subsequently allowing the uniform and numerous formation of alloy carbides.

It is therefore believed that, in addition to the increased strength, the primary contributing factors to the improved rolling-contact fatigue life of the ausformed material is the reduction in size of the carbide particles as well as their more uniform dispersion throughout the structure. From a fatigue standpoint, having small, uniformly dispersed carbides is beneficial because these particles will increase the resistance to wear, while lessening the severity of dislocation pile-ups and hence stress concentrations that accelerate crack initiation or propagation. The current effort in this area is concerned with scaling up the ausforming procedures to produce main shaft engine bearing components ranging up to 210 mm in bore size.

Besides the improvement in rolling-contact fatigue life, ausforming also offers a significant improvement in the fracture toughness characteristics of the bearing steel. This is a material property that is assuming increasingly greater significance in all phases of engineering metallurgy. Because it is pertinent to all bearing material improvements discussed so far, it will be reviewed briefly.

Basically fracture toughness is a measure of the load-carrying capacity of a material in the presence of a crack. It becomes an important consideration in those cases where a bearing component, generally the outer race, is designed to sustain steady-state operating hoop stresses in addition to the normal rolling-contact stresses. The fracture toughness of brittle materials such as M-50 is generally quite low, and consequently allowable hoop stresses must be kept to a minimum. It has been shown, however, that ausformed material has fracture toughness properties as much as five times that of the standard material (ref. 47) and consequently allows considerably more freedom in bearing design.

Fracture toughness considerations are becoming increasingly pertinent to bearing materials technology. This is chiefly because of the increases in the physical size of the bearings and the prevalent design practices in which bearings are used as major structural members.

The method of applying linear elastic fracture mechanics to brittle materials such as bearing steels can be described as follows. The intensity of the tensile stress field in the vicinity of the tip of a sharp crack (such as the bottom of a rolling-contact fatigue spall) can be expressed in terms of a single parameter, the stress intensity factor,  $K$ . In general, the stress intensity factor is found to be proportional to the cross-section unperturbed stress,  $\sigma$ , and a crack dimension,  $a$ , in the form:

$$K = C\sigma\sqrt{a}$$

where  $C$  is a geometric parameter.



At fracture, a critical value of  $K$  will be reached

$$K_{Ic} = C\sigma_f\sqrt{a_c}$$

where  $K_{Ic}$  is called the critical plane-strain intensity factor or the plane-strain fracture toughness, and  $\sigma_f$  is the fracture stress. Thus  $K_{Ic}$  can be used in design against brittle fracture in a manner similar to the use of the yield strength in design against deformation or the ultimate strength in design against rupture. Data relative to  $K_{Ic}$  for bearing steels can be obtained experimentally from simple fracture tests.

#### POWDER METALLURGY

Powder metallurgy techniques are being explored for bearing alloy applications. Specific applications to date are few, although initial reports on NM-100 (ref. 48) indicate that this powder metallurgy product is performing well in certain bearing applications. The major contribution of powder metallurgy to bearing materials technology will be to allow highly alloyed, segregation-prone materials (such as WDC 65) to be produced with a fine grain size and virtually no segregation. Some work is being performed currently on a powder metallurgy produced M-50, again with the objective of alleviating the difficulty of working or machining this material by reducing segregation. Some earlier attempts to apply powder metallurgy techniques to M-50 were not successful (ref. 49), but continual improvements in the processing and consolidation techniques since these first trials warrant additional studies in this area.

There are, of course, other approaches being studied for improving bearing life. The differential hardness principle (ref. 50) is an example of this. Carburized steels also offer a degree of improvement though their main application appears to be in roller bearings. Electron beam irradiation as a method of controlled and limited surface hardening (ref. 51) also has an interesting potential.

#### FUTURE CHALLENGES

In the foreseeable future, the challenge to metallurgists concerned with bearing technology will be even greater than it is today. Engine speeds and sizes are increasing to where DN (bearing bore in mm times inner ring speed in rpm) values of  $3 \times 10^6$  now appear inevitable. Reliability, always a major concern, is magnified by the economic considerations and number of lives involved in one of the new generation of superjets. As hypersonic technology enters the arena, bearing temperatures will begin to increase. However, in view of the 1960 prediction batting average, no numbers will be assigned to these temperature levels at present.

What then are some of the metallurgical activities which will be needed to cope with the above problem? Probably the most exciting of these is the hollow ball. Recent work performed in this area (ref. 52) indicates that important advantages can be realized by reducing ball weight and inertia loading in high speed, large-diameter bearings.

In its present state of development, the hollow ball is still very much in the metallurgist's domain. A great deal must be learned yet regarding the forming of the ball hemispheres, the optimum joining processes, the processing after joining, and perhaps even more basic, the selection of the ball material itself. It is quite conceivable that in order to realize the maximum benefit from a hollow ball (weight reductions of 50 percent or better), present materials such as M-50 or 52100 are inadequate. The problem is that, to achieve a 50-percent weight reduction, wall thicknesses will necessarily be quite thin (OD/ID ratio of 1.2). The poor ductility and low fracture toughness of the M-50 or 52100 will not allow a viable part having this wall thickness under any but the lightest loads. The situation is even more aggravated when an internal defect such as weld flash is present. Under these conditions the ball will fail in bending fatigue long before it reaches its nominal rolling-contact fatigue life.

Consequently it may be necessary to consider other materials, more ductile and less notch sensitive, for hollow ball applications. A fertile area for investigation is the family of work-hardenable materials such as the cobalt- and nickel-base alloys currently being used or being developed for other critical aircraft applications. While these materials initially may not have the hardness levels considered necessary for optimum fatigue life, it has been demonstrated that under the high localized loading present in rolling-element bearings, sufficient additional cold-working is achieved which raises the hardness to acceptable levels. This was shown with a cobalt-base alloy (Haynes Stellite 25) where hardness increases of as much as 12 Rockwell C points were noted in the contact zone (ref. 53). Recently it was also indicated (ref. 54) that at hardness levels of Rockwell C 50 or below, the derating factor may be considerably less than predicted by current formula. If this is indeed the case, then the use of lower hardness, more ductile materials for hollow balls appears even more feasible.

In addition, more emphasis should be placed on modifying existing alloys such as M-50. As mentioned previously, none of the currently used bearing materials were developed specifically for this application. Therefore, while M-50 is a good bearing steel, it is quite possible that a chemistry alteration may improve it both in fatigue life and fracture toughness characteristics.

Advances in metalworking techniques will also need to be evaluated in terms of their potential benefit on bearing material. The area of high



energy rate forming may be a fruitful one for improving the properties of existing steels. Ausforming is only one of several thermomechanical treatments, and certainly other such working methods need to be explored.

Further improvements in melting techniques will be a potential source for improved bearing steels. High-capacity consumable-electrode remelt furnaces, improved and high-capacity electron-beam-melting sources all need to be investigated.

The exact role of inclusions in rolling-contact fatigue needs to be defined. It is quite possible that the present ASTM "JK" system of inclusion rating is totally unrealistic for bearing purposes. It is possible that a bearing-fatigue-oriented inclusion rating system will have to be developed.

Finally materials "tailored" for bearing use and developed on the basis of existing knowledge coupled with realistic tests during their development phase offer hope for achieving the materials necessary to meet the demanding bearing requirements of the next decade.

#### DISCUSSION

**P. M. Ku** (Southwest Research Institute, San Antonio, Texas)

Mr. Bamberger states that approximately 10 percent of the rolling-element bearing failures can be attributed to fatigue, 85 percent to lubricant problems, and 5 percent to other causes. If indeed fatigue is such an infrequent mode of bearing failure, then why the overwhelming emphasis on fatigue life?

It is gathered that Mr. Bamberger is referring to jet engine bearings operating at high temperatures, for which chemical stability and deposit formation of the lubricants probably account for a goodly portion of the "lubricant" problems. If that is the case, then would it be possible that fatigue might be a more important type of failure for many industrial applications with modest operating temperatures?

In general, it would also appear that some of the "lubricant" problems may possibly be "lubrication" or lubricant-metal-atmosphere interaction problems. Here the metallurgists and material scientists can no doubt also play a constructive role.

#### LECTURER'S CLOSURE

Mr. Ku's comments are well taken and are appreciated by the author. The paper was indeed written from the viewpoint of one primarily concerned with jet engine bearing technology. Consequently the distribution of failure causes was quite naturally based on this background and may not necessarily be relevant to other bearing applications.

With regard to the emphasis on fatigue failures, while this type of



failure does represent only a relatively small percentage of the bearing problems observed in aircraft gas turbines, it is nevertheless the only failure mechanism over which the metallurgist can exercise any direct control. Therefore, the primary task of the bearing materials engineer is to assure the development and application of the highest fatigue strength material in these critical engine components. Mr. Ku points out that in non-aircraft applications with less severe operating conditions, fatigue might well play a more prominent role in bearing failures. The author does not have any data available to positively agree with, or take issue with this statement, although it does seem to be a reasonable supposition.

On the other hand, it might be suggested that, in these applications, still other failure modes might be present that would partially replace the "lubrication" failure category so prominently featured as a cause of jet engine problems. For instance, in industrial applications, one might consider the long-time effects of corrosive media—both gaseous and liquid—abrasive materials such as sand, grit, etc., and other similar factors which could precipitate bearing distress long before a fatigue failure would terminate bearing life.

Lastly Mr. Ku seeks a more detailed definition for the generic term "lubrication failure." Basically this area can be divided into two major categories—"mechanical lubrication problems" and "lubricant problems." The former consists of items such as failures of oil supply lines, seal distress or failure, lubricant jet clogging, etc. Obviously these are difficulties which no amount of interdisciplinary rapprochement will effectively eliminate. Conversely "lubricant problems" are frequently associated with fluid-metal-atmosphere reactions which do lend themselves to interdisciplinary solutions. While the importance of these reactions and their relationship to metal failures have long been recognized in other fields, and particularly so in aircraft structural materials applications, only a modicum of work has been done in this important area as related to bearing technology.

For example, it has been shown (ref. 55) that acidic products formed by oil oxidation reduce bearing life, possibly by chemical effects on the surface followed by hydrogen-embrittlement induced failures. It has also been suggested (ref. 56) that the presence of a surface active component in the oil may reduce the surface energy of the metal to such an extent that surface cracks will form more readily under stress conditions than would otherwise be considered moderate. Schatzberg and Felsen (ref. 57) have pointed out the influence of dissolved water in the lubricant on fatigue-failure location and surface alterations.

Beyond this, other lubricant problems can be attributed to rheological effects such as those discussed by Appeldoorn (ref. 58). Still others are due to lubricant viscosity and film thickness effects such as those

shown by Littmann (ref. 59). It is in these areas that an interdisciplinary approach combining the talents of the metallurgist, materials scientist, chemical engineer, surface chemist, and others can function most effectively in increasing bearing life and thereby improving systems reliability.

## REFERENCES

1. SHAW, M. C.; AND MACKS, F.: Analysis and Lubrication of Bearings. McGraw-Hill Book Co., 1949.
2. WILCOCK, D. F.; AND BOOSER, E. R.: Bearing Design and Application. McGraw-Hill Book Co., 1957.
3. ANON.: Proc. of the Air Force-Navy-Industry Propulsion Systems Lubrication Conference. November 1960.
4. PHILIP, T. V.; NEHRENBURG, A. E.; AND STEVEN, G.: A Study of the Metallurgical Properties That Are Necessary for Satisfactory Bearing Performance and the Development of Improved Bearing Alloys for Service up to 1000° F. WADC Tech. Rept. 57-343, 1958.
5. BAMBERGER, E. N.: Bearing Fatigue Investigation. NASA CR-72290, September 1967.
6. BAMBERGER, E. N.; ZARETSKY, E. V.; AND ANDERSON, W. J.: Fatigue Life of 120 mm Bore Ball Bearings at 600° F with Fluorocarbon, Polyphenyl Ether and Synthetic Paraffinic Lubricants. NASA TN D-4850, 1968.
7. KASAK, A.; CHANDHOK, V. K.; AND DULIS, E. J.: Development of New and Useful Elevated Temperature Steels for Aircraft Applications. ASD-TR-61-386, 1962.
8. ANDERSON, W. J.; AND BISSON, E. E.: Choice of Materials is Important for Success with High-Temperature, Dry Bearings. SAE J., vol. 68, 1960, pp. 74-75.
9. ANON.: Jet Liners Get Assist from New Steel Alloy. Western Metalworking, vol. 18, 1960, p. 39.
10. AMATEAU, M. F.: High Temperature Bearing Properties and Cobalt-Base Alloys. Cobalt, December 1960, pp. 27-33.
11. LESCO-BG42: Heat and Corrosion Resistant Steel. Alloy Digest, SS-179, March 1966.
12. CARPENTER, V. S. M.: Hot Work Steel. Alloy Digest, TS-179, May 1966.
13. REGENT VAC-ARC: Bearing Steel. Alloy Digest, SA-188, December 1965.
14. SCOTT, D.: Substitute Alloys As Bearing Materials. Nat. Engrg. Lab., 1967.
15. NIJHAWAN, B. R.: Work on Substitute Alloys at the National Metallurgical Laboratory (India). J. Iron Steel Inst., May 1968.
16. IVANOV, A. G.; AND STEPANYANTS, M. I.: The Effect of Carbon, Chromium and Molybdenum on the (Corrosion and Mechanical) Properties of 1.0% Carbon, 18% Chromium Stainless Steel, 9Kh 18. Sbornik Trudy Tsentr. Navchn.-issled. Inst. Chern. Met., 1967.
17. MONMA, K.; SUTO, H.; AND NAKANO, K. I.: The Effects of the Amount of Residual Carbide and Carbon Content in the Matrix on the Strength of Ball Bearing Steels. Kinzokv Gakkai-Si, November 1967, p. 31.
18. HENRY, R.: Evaluation of New Bearing Alloys. Vanadium-Alloys Steel Co., Handbook, November 1968.
19. VENO, M.; AND NAKAJIMA, H.: Studies on Ball Bearing Steels. Pt. VIII. Behavior of Carbides in Ball Bearing Steels by Electrolytic Isolation. J. Mech. Lab., vol. 11, 1957, pp. 108-113.
20. MORRISON, T. W.; TALLIAN, T.; WALP, H. O.; AND BAILE, G. H.: The Effect of



- Metallurgical Variables on the Fatigue Life of AISI 52100 Steel Ball Bearings. ASLE Trans., vol. 5, no. 2, 1962.
21. AKSOY, A. M.; LILLYS, C. P.; AND JONES, F. G., JR.: Low Alloy Bearing Steel. U.S. Patent 3 306 734, February 28, 1967.
  22. KJERRMAN, B.: Basic Electric Arc Steel Versus Acid Open Hearth Steel for Roller Bearings. ASM Proc. of the First World Metallurgical Congress, 1952.
  23. KISELEV, A. A.; LAPSHOVA, M. P.; AND KUL'KOVA, M. N.: Production of Ball Bearing Steel in Acid Open Hearth Furnace Fired with Natural Gas Mixed with Fule Oil. Stal', January 1958.
  24. SMOLYAKOV, V. F.; KALINNIKOV, E. S.; AND POTAPOV, V. D.: Contamination of Ball Bearing Steel with Refining Slag. Stal', October 1957.
  25. SCOTT, D.: The Effect of Steelmaking, Vacuum Melting, and Casting Techniques on the Life of Rolling Bearings. Proc. Congres Internat. sur les Applications des Techniques du Vide a la Metallurgie, November 1968.
  26. CHURCH, C. P.; KREBS, T. M.; AND ROWAN, J. P.: The Application of Vacuum Degassing to Bearing Steel. J. Metals, vol. 18, no. 1, 1966.
  27. BUZEK, Z.; EMINGER, Z.; HLINENY, J.; AND KLETECKA, Z.: A Comparison of the Effects of Electroslag and Vacuum Arc Remelting on the Properties of Steel. Sbornik Vedeckych Praci, 1967.
  28. PAVPEROVA, I. A.: Quality of ShKh 15 Steel Remelted by Single and Double Electroslag Technique. Stal', no. 9, 1964.
  29. KARPENKO, G. V.; KUSLITSKIY, A. B.; AND BABEY, YU. L.: Effect of Density on Fatigue Strength of Electroslag and Vacuum Remelted Ball-Bearing Steel. Dopovidi Akad. Nauk UKR RSR, no. 8, 1964.
  30. KAGNOVSKI, G. P.; VUL'FOVICH, M. S.; ZABALUEV, YU. I.; AND PAKHOMOV, A. I.: Improved Ball Bearing Steel. Metallurg, no. 11, 1965.
  31. VACHUGOV, G. A.; AND ANTROPOVA, G. A.: Composition and Distribution of Nonmetallic Inclusions in Electroslag Melted Ingot of ShKh 15 Steel. Stal', no. 2, February 1966.
  32. KUSLITSKI, A. B.; MINDYUK, A. K.; RUDENKO, V. P.; AND RYABOV, B. F.: Melting Processes. Soviet Materials Sci., vol. 1, no. 1, 1965.
  33. KLYUEV, M. M.; AND DEDUCHEV, L. A.: Deoxidation in Electroslag Remelting. Avtomat Svarka, no. 5, May 1966.
  34. ALLEN, A. G.: The Electroslag Refining Process. AEI Engrg., vol. 6, no. 4, 1966.
  35. EFIMENKO, YU. M.; MOVCHAN, B. A.; AND TIKHONOVSKI, A. L.: Electron Beam Melting and Refining of ShKh 15 Ball-Bearing Steel. Stal', no. 3, March 1966.
  36. CARTER, T. L.: A Study of Some Factors Affecting Rolling-Contact Fatigue Life. NASA TR R-60, 1960.
  37. MORRISON, T. W.; WALP, H. O.; AND REMORENKO, R. P.: Materials in Rolling Element Bearings for Normal and Elevated (450° F) Temperatures. ASLE Trans., vol. 2, no. 1, 1959.
  38. LYNE, C. M.; AND KASAK, A.: Effect of Sulfur on the Fatigue Behavior of Bearing Steel. ASM Trans., vol. 61, no. 1, 1968.
  39. KOISTINEN, D. P.: The Generation of Residual Compressive Stresses in the Surface Layers of Through Hardening Steel Components by Heat Treatment. ASM Trans., vol. 52, no. 3, 1964.
  40. LIPS, E. M.; AND VAN ZUILEN, H.: Improved Hardening Technique. ASM Metal Progress, August 1954.
  41. SCHMATZ, D. J.; AND ZACKAY, V. F.: Mechanical Properties of Deformed Metastable Austenitic Ultrahigh Strength Steels. Trans. ASM, 1959.
  42. JUSTUSSON, W. M.; AND ZACKAY, V. F.: Engineering Properties of Ausformed Steels. ASM Metal Progress, December 1962.



43. MARSCHALL, C. W.: Draft Literature Survey on Hot Cold Working of Steel. Battelle Memorial Inst. Project AF33 (657) 9139, September 1962.
44. KOPPENAAL, T. J.: The Current Status of Thermomechanical Treatment of Steel in the Soviet Union. *ASM Trans.*, vol. 62, no. 1, 1969.
45. BAMBERGER, E. N.: The Effect of Ausforming on the Rolling Contact Fatigue Life of a Typical Bearing Steel. *Trans. ASME, J. Lub. Tech.*, vol. 89F, no. 1, 1967.
46. BAMBERGER, E. N.: The Production, Testing and Evaluation of Ausformed Ball Bearings. Final Engineering Report, BuWeps Contract NOw-65. 0070-f, June 1966.
47. ZACKAY, V. F. ET AL.: The Strength and Toughness of Dynamically Strain Aged Alloy Steels. VCRL-16363, Univ. of California, October 1965.
48. BRADLEY, E. F.; SPRAGUE, R. A.; AND TUFFIN, W. B.: A New Stainless Steel from Powder. *ASM Metal Progress*, September 1965.
49. HOPKINS, J. M.; AND JOHNSON, J. H.: Method of Producing Improved Bearing Components by Elimination or Control of Fiber Orientation. NASA CR-5J402, 1963.
50. ZARETSKY, E. V.; PARKER, R. J.; AND ANDERSON, W. M.: Effect of Component Differential Hardnesses on Rolling Contact Fatigue and Load Capacity. NASA TN D-2640, 1965.
51. NAZAREMKO, G. T.; AND ZAGAVURA, G. YA.: Electron Beam Treatment as a Method of Work Hardening Steel Surfaces. *Dopovidi Akad. Nauk UKR RSR*, no. 1, January 1963.
52. HARRIS, T. A.: On the Effectiveness of Hollow Balls in High Speed Thrust Bearings. *ASLE Trans.*, vol. 11, 1968.
53. BAUGHMAN, R. A.; AND BAMBERGER, E. N.: Unlubricated High Temperature Bearing Studies. *Trans. ASME, J. Basic Engrg.*, vol. 85D, no. 2, 1965.
54. ANON.: *Machine Design*. vol. 41, January 23, 1969, p. 15.
55. ROUNDS, F. G.: Discussion of Effect of Three Advanced Lubricants on High Temperature Bearing Life by E. N. Bamberger, E. V. Zaretsky, and W. J. Anderson, *Trans. ASME, J. Lub. Tech.*, vol. 92F, no. 1, 1970.
56. LIKTMAN, V. I.; REHBINDER, P. A.; AND KARPENKO, G. V.: Effect of Surface Active Medium on the Deformation of Metals. *Acad. Sci. Publ. House (Moscow)*, 1954.
57. SCHATZBERG, P.; AND FELSEN, I. M.: Influence of Water on Fatigue-Failure Location and Surface Alteration During Rolling Contact Lubrication. *Trans. ASME, J. Lub. Tech.*, vol. 91F, no. 2, 1969.
58. APPELDOORN, J. K.: The Present State of Lubrication and Its Relation to Rheology. *Trans. ASME, J. Lub. Tech.*, vol. 90F, no. 3, 1968.
59. LITTMANN, W. E.; WIDNER, R. L.; WOLFE, J. O.; AND STONER, J. D.: The Role of Lubrication in Propagation of Contact Fatigue Cracks. *Trans. ASME, J. Lub. Tech.*, vol. 90F, no. 1, 1968.

**Page intentionally left blank**

# Effects of Lubricants—General Background

**J. K. APPELDOORN**

**Esso Research and Engineering Company  
Linden, New Jersey**

In concentrated contact, the quality of the oil is frequently dictated by requirements other than friction and wear. In particular, the oil must have good stability: thermal, oxidative, and hydrolytic. For the lubrication process itself, viscosity is the only physical property of importance, but chemical properties are even more important. The selection of these depends on whether the limiting failure mode is scuffing, gradual wear, or fatigue pitting.

**A** LUBRICANT in a rolling-contact bearing or a gear may be called on to perform many functions. The most obvious is that it should lubricate; that is, it should reduce friction and wear. But it may also serve as a coolant, conducting away the heat generated in the bearing and keeping the temperature at a workable level. It may be a sealant, serving as a grease does to keep dirt particles out of the running track. And it may be a corrosion inhibitor, protecting the metal surfaces from rust and chemical attack.

The lubricant must also meet many negative requirements. Its viscosity must not change too much with changes in temperature. It must not foam, freeze, separate into two phases, evaporate, corrode metal surfaces, swell seals, form sludge, or emulsify with water. Most important, it must not oxidize or undergo thermal decomposition. Indeed, this last requirement is frequently the limiting one in formulating an oil. In many applications, almost any oil is a satisfactory lubricant if it can simply exist, that is, if it can remain relatively unchanged under the conditions of its environment, no matter how extreme.

The jet engine lubricants are a good example of this. Jet engines operate at high speed and require a low-viscosity oil. The first jet engines used a light petroleum oil, but it was soon found that these oils would evaporate excessively under the high-temperature conditions of the hot gas turbine. Synthetic diesters were then substituted. These lubricants were of a single molecular species and therefore had no light ends to



boil away. They could be constructed to give the optimum combination of low viscosity and low volatility by making them long straight-chain molecules with just enough branching to ensure a low freezing point. Oxidation stability was provided by new kinds of inhibitors.

As new jet engines appeared, internal temperatures increased, and more vigorous demands were made on the lubricant. Oxidation inhibitors were no longer a complete answer. The oil molecules themselves had to be made more resistant to attack. Inasmuch as this attack is primarily at the C-H bonds, the weakest of these bonds (those located  $\beta$  to the ester group) had to be eliminated. These "no  $\beta$ -hydrogen" esters provided a third generation jet engine oil. Today the main line of research is to get more oxidation stability so that still higher temperatures can be met (ref. 1).

This same situation exists in many other lubricant fields. The excitement when a new synthetic oil is introduced is almost never because of an improved lubricating ability but because of other properties, particularly a superior oxidative and thermal stability (ref. 2). Such lubricants as the silicones, polyphenyl ethers, halocarbons, and the new perfluoro-polyethers all have serious drawbacks in price or performance; several are only marginal in lubricating ability. They are used because they have the overriding advantage of excellent stability at high temperatures (ref. 3).

This point has been made first because it is sometimes forgotten when discussing the lubrication mechanism itself. The choice of an oil is frequently dictated by requirements other than lubricating ability. The choice of greases or solid-film lubricants provides other examples. Greases are oils that have been thickened with soap or finely divided solids to give a structure that has a finite yield stress. Under the shearing forces between the moving surfaces, the structure is pseudoplastic and the viscosity approaches that of the base oil. But outside this shear zone the grease stays as a solid, preventing oil from leaking out or dirt from getting in. The purpose of using a grease rather than an oil is not because it has better lubricating ability, but because it is a sealant.

Solid-film lubricants are usually lamellar-layer materials that are applied to bearing surfaces intended for operation where liquid lubricants are not applicable. In outer space, for example, liquid lubricants are inadequate because they will evaporate in the hard vacuum and will be degraded by radiation. In food-processing machinery, liquid lubricants are objectionable because they may cause contamination. Solid lubricants have neither of these limitations and so are widely used (ref. 4). In these applications, however, it should again be noted that the chief reason for use is something other than superior lubricating ability.

Turning now to the lubricating process itself, which lubricant properties are important? To answer this question, it is necessary to know what

the failure mode is likely to be, for different failure modes require different cures.

For this discussion, the failure modes will be grouped into three categories: wear, scuffing, and fatigue. Wear is defined as a gradual and constant attrition of the contact surfaces by abrasion, corrosion, or plowing and may range from negligible to rapid. In contrast, scuffing is a gross failure of the surface films, and the transition from wear to scuffing is often dramatic. Fatigue is the pitting of the surface from repeated contact stressing.

This grouping of failure modes into wear, scuffing, and fatigue emphasizes the unique character of scuffing and fatigue as contrasted to normal wear. Many previous authors have used this grouping, though their nomenclature may differ. Benedict (ref. 5) divides the failure modes of gears into these three areas but calls them instead "wear, scoring, and pitting." Godfrey (ref. 6) suggests in place of scuffing to use the term "load-carrying capacity," this being the load at which the lubricant film fails (i.e., above which scuffing occurs). Lipson (ref. 7) uses the terms "pitting and spalling" in place of fatigue; "adhesive wear" in place of scuffing; and subdivides wear into abrasion, cavitation, and corrosion. Archard and Hirst (refs. 8 and 9) and Lancaster (ref. 10) use the terms "mild wear," which leaves the surfaces smooth and the surface films intact, and "severe wear" which leaves the surfaces torn. (This terminology has the objection of implying that mild wear is not serious, for "serious mild wear" seems like a contradiction in terms, whereas actually it can be extremely rapid.) The distinction between scuffing and all other kinds of wear is also implicit in such common usage terms as transition temperature, antiwear additives vs EP additives, or the four-ball wear tester vs the four-ball EP machine.

#### WEAR

Wear, then, is a gradual process. Also, regardless of the actual mechanism by which it occurs, it can be thought of as a thin-film phenomenon. The mechanism may involve corrosion, erosion, plowing, etc., or a combination of these (ref. 11); but in any case the wear can be ameliorated if thicker oil films are present. Thus higher speeds are beneficial because they promote more full-fluid-film lubrication. In rolling-contact bearings especially, where rolling speeds are high (and the amount of sliding is relatively low), wear is not generally a serious problem.

Thicker films can also be obtained by using an oil of higher viscosity. Indeed, the instinctive way to get out of any trouble is first to choose a heavier oil. For this reason, most oils are specified by viscosity grades. Sometimes viscosity is dictated by other factors. A thick oil film in a ball bearing may cause skidding instead of rolling. Or the friction losses from viscous drag may militate against heavy oils. Should this method



be impractical or inadequate, the chemical make-up of the oil must be changed, usually by employing special additives.

Fatty acids have long been known as good antiwear additives. They have been studied extensively and are used commercially to some degree (ref. 12). These materials adsorb or react at the rubbing surface to give a film (perhaps only one molecular layer thick) that lessens adhesion (ref. 13), corrosion (ref. 14), and abrasion (ref. 15). Other surface active agents such as alcohols, amines, esters, etc., are similarly effective to a greater or lesser extent (ref. 16). Adsorption will be discussed at greater length later in this symposium.

The role of these additives is fairly straightforward, but it is more complex than a simple adsorption from an inert solvent. Askwith et al. (ref. 17) has shown that a  $C_{16}$  fatty acid is most effective when adsorbed from a  $C_{16}$  hydrocarbon solution and is less effective if a  $C_{14}$  or  $C_{18}$  hydrocarbon is used. Fein (ref. 18) has similarly noted that oleic acid is very effective in *n*-hexadecane ( $C_{16}$ ) but much less effective in squalane ( $C_{30}$ ). Apparently the film contains some of the solvent hydrocarbon, or else the rate of adsorption is considerably dependent on the base stock used (ref. 19).

Another class of additives—of which tricresylphosphate (TCP) is a prominent example—prevents wear by reacting with the surface rather than merely being chemisorbed. They form what are believed to be inorganic films such as iron phosphate which give the necessary protection. These additives are sometimes classified as mild EP agents and will be discussed more thoroughly in the next section under scuffing.

Even the composition of the base stock without additives can have a pronounced effect on wear. For example, condensed-ring aromatics when added to aliphatic hydrocarbons give considerably better antiwear properties than either component alone (ref. 20), a phenomenon that is not at all understood. Finally the surrounding atmosphere can also affect wear appreciably. Both oxygen and moisture tend to increase corrosive wear significantly (refs. 14, 21, and 22) and generally irrespective of the base stock used.

Some specific kinds of wear require specific measures. Abrasion may result if hard particles are suspended in the oil and become trapped between the rubbing surfaces. Super-filtration on a continuous basis may be required here (ref. 23). Fretting corrosion is the result of the combination of insufficient lubricant and a vibratory motion. Metal fragments are "plucked" from the surfaces by a frictional process and then rapidly oxidize to cause further damage by abrasion. Although some antiwear additives are quite effective, the best solution here is to use a solid lubricant to prevent the metal-to-metal contact in the first place (ref. 24).

In summary, wear is primarily a thin-film phenomenon that can be



alleviated by using higher speeds, more viscous oils, or surface-active additives.

### SCUFFING

Scuffing is an entirely different phenomenon from ordinary wear and is the result of an overall failure of the surface films. At very slow speeds or in unlubricated sliding, scuffing may merely appear as another form of wear. But in gears or bearings, a scuffing failure is usually catastrophic. Indeed, for gear oils, the most important property is that it prevents scuffing.

Scuffing differs from wear in a second important way—the response to higher speed. Wear tends to become less severe at higher speeds because the film thickness is greater. (Landen (ref. 25) has recently reported on the wear problem in gears operating under slow-speed conditions.) Scuffing, however, becomes more severe at higher speeds (ref. 26); that is, the load-carrying capacity of gears decreases as the speed increases. Scuffing, therefore, is a high-temperature phenomenon as pointed out by Blok 30 years ago (ref. 27). These high surface temperatures are a result of frictional heating either in the oil film or by actual metallic contact. The “flash” temperature at the surface can be calculated from the friction coefficient (ref. 28), and it has been shown that oils of a single chemical type will all fail at the same surface temperature (ref. 29). This temperature probably corresponds to a desorption or softening of the surface films (ref. 3).

Viscosity can be of considerable help in alleviating scuffing. Thicker oils mean thicker films and less metal-to-metal friction. However, viscosity is not as effective as it is in reducing wear because the problem is not how to get a thicker oil film but rather how to reduce frictional heating. And excessive friction can be developed in the oil film even though the metal-to-metal friction is reduced. Nevertheless, gear oils are classified by viscosity grade, the higher viscosities intended for more severe service (ref. 30).

For still more severe service, extreme-pressure (EP) additives are required. Their action is quite different from that of the antiwear additives. Antiwear additives merely adsorb—physically or chemically—on the surface; and when the temperature gets too high, they desorb (ref. 31). EP additives, on the other hand, take advantage of the high temperatures by decomposing and reacting with the metal surface to form a new kind of film (ref. 32), probably an inorganic compound (refs. 33 and 34). The film plays a dual role; it prevents welding of freshly-exposed surfaces, and it provides an easily-smearable layer that reduces friction and therefore heat. It is thus both a preventive and a cure.

However, the kind of additives to be used depends on the kind of service. For high-speed low-torque operation, a lead-soap active-sulfur additive is best (ref. 35). (It is interesting to note that 1969 marks the

hundredth anniversary of the first EP additive (ref. 36)—a lead-soap active-sulfur type.) The important requirement here is that the additive reacts very rapidly with any fresh metal surface; the resultant film does not wear away rapidly because the loads are relatively low.

The same additive in low-speed high-torque service is quite unsatisfactory, for it is too reactive. Excessive amounts of the surface films are formed and these, having low shear strengths, are rapidly worn away (ref. 37). EP additives, therefore, can be thought of as materials that give a controlled amount of corrosive wear in order to prevent scuffing (ref. 38). If the amount of corrosive wear is not controllable, the cure may be worse than the disease.

Considering that EP action is essentially chemical in nature, it is somewhat startling to find that current knowledge of the actual reaction involved is still rudimentary. It is known that the EP activity is related to chemical reactivity (refs. 39 to 43) and that EP films contain atoms of the EP additives (refs. 44 to 47). But most of the identification of the composition of EP films is circumstantial, based on reactions with iron powder (ref. 48) or iron wire (ref. 49), or deduced from the effect of changes in the chemical composition of the oil. There has been a long-standing belief that the film formed by TCP is mostly iron phosphide (ref. 50), or that the film formed from sulfur compounds is mostly iron sulfide (ref. 51). Godfrey has helped refute these beliefs by his electron diffraction studies which showed only iron phosphate present in the first case (ref. 52) and mostly iron oxide in the second (ref. 53). It is also significant that he found the iron phosphate pattern only when TCP was used as the whole lubricant and not when he studied TCP solutions in oil. A better understanding of surface reactions is one of the most necessary and most challenging problems in the entire field of lubrication. This will also be discussed more thoroughly later in this symposium.

In spite of the empirical nature of EP-oil formulation, it should not be thought that EP oils are a crutch, something to take care of bad design or inadvertent overloading. The hypoid gear, found in the drive train of most cars, is a case in point. It represents an extreme example of concentrated contact and high speed. At the pitch line the motion is almost pure sliding and the sliding speeds can reach 2000 ft/min. Yet this gear routinely transmits 200 hp, withstands occasional drag-strip-like starts, and operates without attention for the life of the car. The gear is designed in the knowledge that it will be operated with an EP oil. The oil becomes a part of the design itself—an interdisciplinary approach that should be more widely followed.

#### FATIGUE

Fatigue pitting is an "old-age" failure. If the gear or bearing functions satisfactorily without scuffing or undue wear, it eventually fails from



fatigue, much as a person who, surviving sickness and accident, eventually dies from the aging process itself. Fatigue is the result of repeated stressing that causes the formation of microcracks. These grow until a bit of metal is dislodged, leaving a characteristic pit.

The role of the lubricant in influencing fatigue is least understood of all. In the case of scuffing and wear, the explanations may be lacking, but the data are clear. There is no question that oil properties are important, and there are well recognized ways to impart better scuff resistance or antiwear character into an oil. In fatigue, however, there is no agreement whether any lubricant property is important, or if so, whether it is helpful or harmful.

According to the AFBMA-ASA standards in selecting the proper size bearing, the critical factor is load; the type of lubricant used is not considered at all. Justification for ignoring lubricant properties is that fatigue is a stress phenomenon, and the stress distribution in lubricated elasto-hydrodynamic contact is nearly identical to dry Hertzian contact (ref. 54). Since the lubricant does not change the stress pattern, it would not be expected to influence fatigue life. Besides, the maximum stress is below the surface where the lubricant cannot penetrate.

A modification of the AFBMA-ASA procedure has recently been suggested (refs. 55 to 57) which takes into account the oil film thickness. Using elastohydrodynamic theory, the minimum film thickness is calculated and compared to the surface roughness. If the film is four times as thick as the average asperity height (i.e., if asperity contact is extremely rare), then the bearing rating life is considered to be at least twice the catalog value. This treatment emphasizes the importance of viscosity, for viscosity is the only lubricant property significantly affecting film thickness. (Pressure-viscosity coefficient,  $\alpha$ , also enters the equation, but  $\alpha$  is largely a function of viscosity. For oils of the same viscosity,  $\alpha$  rarely differs by more than a factor of two. The effect of  $\alpha$  alone will be discussed later on.) According to this view, oils of the same viscosity but of different chemical types would be predicted to give the same fatigue life provided there was a full EHD film.

This conclusion has not been verified experimentally. Fatigue life data under full EHD conditions are not routinely obtained or are obtained only with a single class of oils. The effect of molecular type under these conditions has not been determined.

When higher loads are used (700 000 to 1 200 000 psi Hertz), substantial differences in fatigue life have been found among oils of different chemical type (refs. 58 to 61) regardless of their viscosity. Rounds (ref. 61) found that viscosity could increase fatigue life by only a factor of two, whereas chemical type could cause a change by a factor of 10. Moreover, with some classes of oils, the use of a higher viscosity actually resulted in a decreased life.



Rounds has also pointed out (ref. 62) that additives can greatly affect life. The effect is quite complicated as shown in table 1. Some of the additives increased life, some decreased life, some had little effect. In the majority of cases there was an optimum additive concentration, usually 1 to 2 percent, beyond which fatigue life again decreased.

TABLE 1.—*Additives Affect Fatigue Life\**

Additive	Relative fatigue life (Base oil = 100)	
	2% conc.	10% conc.
Oleic acid	200	45
Oleic 1,3-propylene diamine	370	350
Sulfurized sperm oil	270	60
TCP	80	90
Chlorinated wax	35	10

\* See reference 62.

The effect of oil type, as well as operating conditions, was especially emphasized in a recent CRC study (ref. 63). Five laboratories ran fatigue tests on three identical oils. The Hertz stress ranged from 400 000 to 1 200 000 psi. Most of the laboratories rated them in the order: methylchlorophenyl silicone (best), mineral oil, diester (worst), i.e., in reverse order from chemical reactivity. One laboratory rated the silicone worst. This laboratory used the highest load and smallest oil supply (i.e., the conditions were those most likely to give boundary lubrication failure), and the oil that failed was the one having the poorest boundary lubrication properties. The importance of chemical properties on fatigue is also shown by the effect of dissolved water which reduces fatigue life significantly (refs. 64 and 65).

In order to reconcile these data, it appears necessary to postulate that fatigue pitting proceeds by more than one mechanism. Under textbook conditions where there is a full elastohydrodynamic film, it may be that lubricant type or viscosity has no effect, although this has never been demonstrated experimentally. Under less ideal conditions where there is some metal-to-metal contact, chemically active oils apparently lower the fatigue life. This suggests that there is a corrosion aspect to rolling-contact fatigue just as there is to bending fatigue. Under still less favorable conditions, chemical activity appears to be beneficial, and the same oils that reduce friction and wear also reduce fatigue. This kind of fatigue appears to be initiated by the poor lubrication itself.

An interesting side result of this kind of thinking is that if oils can behave in this way so can steels. That is, if the best oil under one set of conditions is not the best oil under all conditions, then the best steel under one set of conditions is probably not the best steel under all conditions. Some data have been published to this effect (ref. 62). This means that thought should be given to designing the oil and the bearing as a unit and not each one separately.

#### OTHER FACTORS

Several other lubricant properties have been suggested as being important in the lubrication process, but these are largely unconfirmed. Three will be discussed here: pressure-viscosity coefficient, wettability, and viscoelasticity (relaxation time). Although it may seem reasonable that each of these properties should have some effect, there is little proof in their favor and actually considerable evidence against them.

#### Pressure-Viscosity Coefficient

At the high pressures existing in the contact area, the viscosity of the oil is many-fold higher than it is at atmospheric pressure. This has led to the measurement of the viscosity-pressure characteristics of a large number of mineral oils (ref. 66), pure hydrocarbons (ref. 67), and synthetics.\* Although the highest pressures used in these experiments (150 000 psi) do not approach the Hertzian contact pressures (upwards of 300 000 psi), the data do give a first approximation of the pressure-viscosity coefficient,  $\alpha$ . The flattest curves belong to the dimethyl silicones and the steepest to the halocarbons and polyphenylethers, mineral oils being intermediate.

It would seem reasonable that an oil having a high viscosity-pressure coefficient would give a thicker oil film. Indeed, all derivations of elastohydrodynamic film thickness include the term  $\alpha$ , usually in the form:

$$h = k[\alpha\mu_0]^{0.7}$$

where  $h$  = minimum film thickness and  $\mu_0$  = viscosity at atmospheric pressure. Harris (ref. 57) actually presents a chart that predicts a naphthenic mineral oil will give a 75-percent thicker film than a diester, because of its greater  $\alpha$ .

Such confidence in elastohydrodynamic theory is not yet warranted because there is no experimental evidence to support it. To the contrary, Tallian et al. (ref. 56) have shown that film thickness is better predicted

\* Fresco, G. P.; Klaus, E. E.; and Tewksbury, E. J.: Measurement and Prediction of Viscosity-Pressure Characteristics of Liquids. *Trans. ASME, J. Lub. Tech.* (to be published).

if  $\alpha$  is omitted rather than included. His data are given in table 2 which shows a series of oils operated under such conditions that the actual film thickness was 6.8  $\mu$ in. The calculated film thickness using  $\alpha$  is seen to give worse agreement than that not using  $\alpha$ . More recently, Kannel et al. (ref. 68) also found that better correlation is obtained when  $\alpha$  is omitted from the equations.

Not only does  $\alpha$  fail to affect film thickness as predicted, it also has little effect on fatigue life (which presumably is dependent on film thickness). Although a correlation between  $\alpha$  and the fatigue life of four oils has been published (ref. 69), an error has since been found in one of the values. The oil that gave the longest fatigue life actually has the smallest  $\alpha$  (not the largest as reported). This oil is a dimethyl silicone, a class that has been found by many other investigators to be excellent in fatigue (refs. 58, 60, and 63) in spite of the flat viscosity-pressure coefficients.

This failure to demonstrate an advantage for high  $\alpha$ 's—either in thicker films or in those performance factors supposed to benefit from thicker films—is a serious weakness in elastohydrodynamic theory. It should be resolved as soon as possible. So far, there has not been much speculation on this point because the amount of data is too sparse. When more data are available, it seems likely that not only the importance of pressure-viscosity coefficient but also the importance of viscosity itself will need to be re-evaluated vis-a-vis the importance of chemical structure, as pointed out earlier in this paper.

#### Wettability

Wettability has been studied extensively with respect to lubrication, particularly by Zisman's group at the Naval Research Laboratory (ref. 70). It has three aspects: adhesion, wetting, and spreading. All of these are related to the contact angle,  $\theta$ , between a drop of oil and the surface.

With regard to adhesion, it can be stated that all oils will adhere to

TABLE 2.—*Pressure-Viscosity Coefficient Does Not Predict Film Thickness\**

Oil	Vis/100° F, cs	Vis-press coeff., kpsi <sup>-1</sup>	Film thickness, $\mu$ in.		
			Exptl.	Using $\alpha$	Not using $\alpha$
Mineral oil—decalin	7.9	---	6.8	4.3	5.1
" "	11.9	---	6.8	5.3	6.4
" "	15.4	.25	6.8	2.9	6.7
Ester	28.4	.16	6.8	3.7	5.9
Mineral oil	79.0	.25	6.8	5.6	6.3

\* See reference 56.



all surfaces. Only if the contact angle is  $180^\circ$  will adhesion be zero as can be seen from the equation for the work of adhesion. This equation, for a non-spreading oil, reduces to:

$$W = \gamma[1 + \cos \theta]$$

where  $\gamma$  is the surface tension. Since the highest contact angle observed experimentally (mercury on stainless steel) is only  $154^\circ$ ,  $W$  is always positive and adhesion always occurs. Thus, the rheologist's assumption of "no slip at the wall" is completely valid, and the thickness of the lubricating oil film is unaffected by wettability.

The second aspect, wetting, depends on whether the contact angle is greater or less than  $90^\circ$ . Thus, n-hexadecane wets PTFE ( $\theta = 45^\circ$ ) and rises in a PTFE capillary; water wets glass ( $\theta = 0^\circ$ ) and rises in a glass capillary; mercury does not wet glass ( $\theta = 132^\circ$ ) and is depressed in a glass capillary. Wetting, so defined, does not appear to be important in lubrication.

Spreading occurs when  $\theta = 0^\circ$ . In this case, a drop of oil will spread to a monomolecular layer, given sufficient surface. Zisman has shown that non-spreading oils are important in certain applications such as the jewel bearings in watches (ref. 71). Being non-spreading, the oil stays at the pivot. Spreading can also be prevented by laying down a contaminating film barrier beyond which the oil will not spread. Seal leakage is minimized in this way (ref. 72).

Spreading can sometimes be a requisite in supplying a lubricant to the bearing, particularly in zero-gravity fields. Also, it has recently been shown that spreading is important in certain gyro bearings. It was found that bearings having surfaces cleaned (and contaminated) with a household detergent gave earlier failure than those with uncontaminated surfaces (ref. 73). The oil supply in this case was a few milligrams of lubricant contained in a porous retainer.

These cases, however, are merely examples of the earlier statement that the most important property of a lubricant is simply to be at the rubbing surface; they do not pertain to the lubricating process itself. In fact, the experimental evidence is that wettability has little to do with lubrication. Zisman and his coworkers have studied the problem exhaustively, and an important result of their work has been "to dispel the old widely held belief that a good lubricant wets better and adheres more to a bearing than a poor lubricant." They point out that TCP is a good boundary lubricant but does not spread, whereas the dimethyl silicones are poor boundary lubricants but spread exceptionally well on all metals (ref. 3). Thus wettability is neither a necessary nor a sufficient requirement for good lubrication.

There is ample supporting evidence of this. A fatty acid film on the surface of a metal will lower its surface energy so much that petroleum

oils will not spread. Yet such fatty acid films do not make for poorer lubrication, as is well known. In the example of the starved gyro bearing quoted earlier, it was also found that non-spreading was not the sole criterion. A contaminating layer of household detergent gave trouble, but an equally contaminating layer of oleic acid did not.

To some extent this question is academic. Bearings and gears are usually metallic, and metals are high energy surfaces having surface free energies upwards of 500 ergs/sq cm. Even nonmetallic materials such as nylon have high surface energies compared to lubricants. Therefore, all ordinary lubricants will spread on all surfaces unless the surface is contaminated with a low-energy film. If this film were to interfere with the lubrication process, it would quickly wear off, exposing the underlying metal which would then be wetted by the lubricant. If the film were to interfere with the transport of the lubricant to the contact area, it is better to ensure that the contamination does not occur rather than try to improve the wettability properties of the lubricant.

#### VISCOELASTICITY

A century ago Clerk-Maxwell showed from theoretical reasoning that an oil subjected to a very high shear stress would no longer behave as an ideal liquid, but would exhibit some solid-like behavior. Specifically the oil would have a measurable modulus of elasticity,  $G$ . Liquids behaving in this fashion are termed viscoelastic, for they behave both as liquids (viscous) and as solids (elastic) depending on the rate at which shear is applied.

Many different test methods have been used to measure viscoelasticity (ref. 74), and it has been shown that there is good agreement among them (ref. 75). A number of oils have been reasonably well characterized with regard to their viscosity and viscoelastic properties over a wide range of shear stress. Polymer solutions begin to show viscoelastic effects at around  $10^3$  dynes/cm<sup>2</sup> (refs. 76 and 77), polysiloxanes at about  $10^6$  (ref. 78), and mineral oils at about  $5 \times 10^7$  (ref. 79), independent of temperature.

However, as predicted by theory (refs. 80 and 81), it has also been shown that the onset of non-Newtonian behavior (viscosity decrease) coincides with the onset of viscoelastic behavior. Elastic behavior therefore becomes appreciable only when there has been severe shear-thinning. So far, experiments have shown that the lower viscosity from non-Newtonian behavior predominates over viscoelastic effects (refs. 82 and 83).

To some extent the nature of viscoelasticity has been misunderstood and taken to mean that when an oil becomes viscoelastic it becomes hard and unyielding (ref. 84). The idea has been advanced that if the



relaxation time  $T$  ( $T = \mu/G$ ) is longer than the time of loading (as between two gear teeth), the oil will give a thicker film.

Actually Maxwell's theory states that a viscoelastic oil will find it easier to deform elastically than to flow viscously. The oil film, therefore, will be even thinner than if no elastic behavior had taken place. This has been demonstrated mathematically for several geometries (refs. 85 to 87). The experimental findings that gears sometimes improve in load-carrying capacity at very high speeds (ref. 88) is most probably due to the beginning of elastohydrodynamic lubrication rather than to viscoelasticity.

If viscoelasticity were important, polymer-thickened oils should be the first to show beneficial effects because their relaxation times are much longer than ordinary liquids. Yet polymer-thickened oils generally give poorer performance in concentrated contacts (ref. 89) and have shorter fatigue life (ref. 60)—a reflection of their shear-thinning viscosity.

It is concluded that the benefits of viscoelasticity, like those of pressure-viscosity coefficient, still await unequivocal experimental proof.

### CONCLUSIONS

The important oil properties for lubrication in concentrated contacts frequently are dictated by other considerations than the lubrication process itself. Also they depend on the type of service. Viscosity is the only physical property that is important regardless of the failure mode, but it is usually not as important as chemical properties. For scuffing, an active chemical is needed to maintain a good surface film. For non-scuffing wear, a protective film of a surface-active additive is sufficient. For fatigue pitting, active chemicals may be harmful in some cases and helpful in others. There is little experimental evidence of the importance of pressure-viscosity coefficient, wettability, or viscoelasticity.

### DISCUSSIONS

**E. E. Klaus** (The Pennsylvania State University, University Park, Pennsylvania)

Mr. Appeldoorn has presented a compelling case for the more careful consideration of the chemical properties of lubricants in the critical study of lubricants in concentrated contacts. This point is certainly well taken and should not generate major disagreement. The role of physical properties in the lubrication of concentrated contacts has been relegated to that of viscosity which then becomes a target for doubt concerning its role in lubrication. In general, there is little evidence in the readily available literature to show the clear-cut relationship of physical properties of lubricants to lubrication failure. However, there appears to be evidence that failures in concentrated contact lubrication may be in-



fluenced by such physical properties as viscosity, volatility, gas solubility, foaming, thermal conductivity, specific heat, and surface wetting.

Measurement of viscosity as a function of temperature, pressure, and shear is now well established for any single variable. However, the combination of any two of these variables under extreme conditions results in a research problem rather than a routine measurement. The idea of measuring the viscosity of the lubricating fluid under the transient conditions encountered in concentrated contacts is relatively new even as a research approach. These efforts appear to suffer at the present time from the problem of too many unknowns. That is, geometry, environmental conditions, and fluid viscosity under the specific conditions are all subject to considerable conjecture. To obtain one value, certain assumptions must be made in the other two areas. As a result, there are substantial differences in values and opinions reflecting the discipline from which one approaches this problem. The complexity of the problem experimentally should not diminish the importance of viscosity under the actual conditions in concentrated contact.

It has been known for some time that gear boxes cause severe degradation of polymer-containing lubricants. That is, these lubricants undergo a permanent viscosity loss of the type caused by flow under severe turbulence and/or cavitation. There is recent evidence that this behavior also applies to ball and roller bearings. A lubricant can undergo the same amount of polymer degradation in 10 to 20 minutes in a bearing as that resulting from 5000 cycles through a hydraulic pump and relief valve. This evidence may supply some important information on the environment that exists and the physical beating the lubricant takes in concentrated contacts.

Fluid volatility has been shown to be an important factor in wear in hydraulic pumps and certain wear testers. This property would also be expected to play a role in the lubrication of concentrated contacts. In instrument bearing testing as well as in a gear type hydraulic pump, wear appears to increase when volatility is increased without an apparent change in viscosity.

Gas solubility and foaming properties appear to be related to lubrication in concentrated contacts. The evidence of turbulent flow in this type of lubrication provides a source of agitation to create foam from a supersaturated solution. The generation of a substantial vapor pressure (more than 50 mm Hg) will also cause the fluid to lose previously dissolved gas. The incidence of foaming can result in a temperature increase at the conjunction since the foam will not carry away as much heat as the liquid. In some gear boxes, relationships have been demonstrated between overheating and foaming, or fluid volatility.

Thermal conductivity and specific heat are physical properties that vary considerably over the range of organic liquids considered for use as

functional fluids and lubricants. Such differences would be expected to influence surface and fluid temperatures in the conjunctions under concentrated contacts. It should be noted that the differences in these thermal properties are relatively minor among typical mineral oil lubricants of a given family. However, the entire range of lubricants must be considered in such a survey.

In the area of surface wetting, the author does refer to the case where a non-wetting film on the metal surface of an instrument bearing is considered to be the reason for premature bearing failure. This well documented case is then swept under the rug by the statement that "TCP is a good boundary lubricant but does not spread, whereas dimethyl silicones are poor boundary lubricants but spread exceptionally well on all metals. Thus, wettability is neither necessary nor a sufficient requirement for good lubrication."

These same data do, in fact, prove that the surface wetting characteristics as a physical property play a very significant role in lubrication. In the case of the non-wetting instrument bearing, the lack of a continuous liquid film affects the type of lubrication that is involved in the bearing operation. Insufficient oil in the conjunction for whatever reason causes a transition from elastohydrodynamic to boundary lubrication with dramatic results. In the case of the TCP-silicone comparison, it has been shown that in the absence of chemically reactive groups the material with the lowest surface tension tends to be physically adsorbed on the solid surfaces. In such a contest between TCP and silicone, it can be shown that the TCP does not appear in appreciable concentrations on the solid surface when these two materials are tested as a binary mixture or both are put in a mineral oil base stock. In both cases the predominant surface film appears to be the silicone.

To summarize, there appear to be a number of examples where physical properties of the lubricant play a significant role in the resultant lubrication. If we are to achieve a better understanding of the fundamentals of lubrication of concentrated contacts, it may be desirable to evaluate these physical effects much more critically rather than dismiss them entirely.

**A. Dyson (Shell Research Limited, Thornton Research Center, Chester, England)**

I have one question for Mr. Appeldoorn and one observation to contribute on a statement made in his paper. First, I wish to ask a question on the statement under the heading "Viscoelasticity" that "the experimental findings that gears sometimes improve in load-carrying capacity at very high speeds is most probably due to the beginning of elastohydrodynamic lubrication . . ." Does the author imply that elastohydrodynamic lubrication has not yet begun at speeds lower than that corresponding to the minimum in the load-carrying capacity? If so, this



is a most remarkable statement, completely at variance with the general consensus of opinion on the subject, and it would be interesting to learn the grounds on which the author bases this opinion.

I also wish to challenge the author's statement under the heading "Pressure-Viscosity Coefficient" that there is no evidence that the pressure-viscosity coefficient affects the elastohydrodynamic film thickness. I wish to present some results obtained by my colleague, A. R. Wilson, who estimated the film thickness from measurements of the electrical capacity between the discs in a disc machine. The experimental details were as given in reference 90, and the oil was an LVI mineral oil, oil D of reference 90. The discs were in pure rolling motion.

In figure 1a, the film thicknesses are plotted against  $\eta_0 \bar{U}$ , where  $\eta_0$  is the viscosity of the oil at atmospheric pressure and at the temperatures of the discs as indicated by thermocouples embedded in them, and  $\bar{U}$  is the rolling speed. The points are seen to lie on two curves, corresponding to the two different disc temperatures used, 44° C and 85° C. The lines are parallel on the log-log scale, the ratio between the two sets of results being approximately 1.27. In figure 1b the same experimental results are plotted against  $\eta_0 \bar{U} \alpha$ , where  $\alpha$  is the pressure-viscosity coefficient, and the points obtained at the two different temperatures fall on one line. The relevant properties of the lubricant at the two different temperatures are given in table 3. The theoretical ratio  $[(2.98 \times 10^{-9}) / (2.17 \times 10^{-9})]^{0.7} = 1.25$ , which is very close to the value of 1.27, read from the vertical displacement between the two lines in figure 1a.

This is only one example of the comparable results that we have obtained with many different lubricants and that demonstrate the importance of the pressure coefficient of viscosity. The very close agreement between theory and experiment demonstrated in reference 90 would be difficult to understand if the pressure coefficient of viscosity were of no importance in the determination of the film thickness, since this pressure coefficient occurs in the theoretical expression used to test the experimental results.

It seems possible that the apparent lack of effect of the pressure coefficient of viscosity on the film thickness, observed in the experiments quoted by Mr. Appeldoorn, was attributable to experimental scatter.

**H. Naylor (Shell Research Limited, Thornton Research Center, Chester, England)**

Mr. Dyson has already covered several points arising from Mr. Appeldoorn's paper, in particular those concerned with the viscoelastic behavior of lubricants and with the significance of the pressure coefficient of viscosity in elastohydrodynamic lubrication. I should like to add further comment on the latter in connection with results, quoted in the presentation of the lecture, which claimed anomalous behavior for glycerol on this basis.



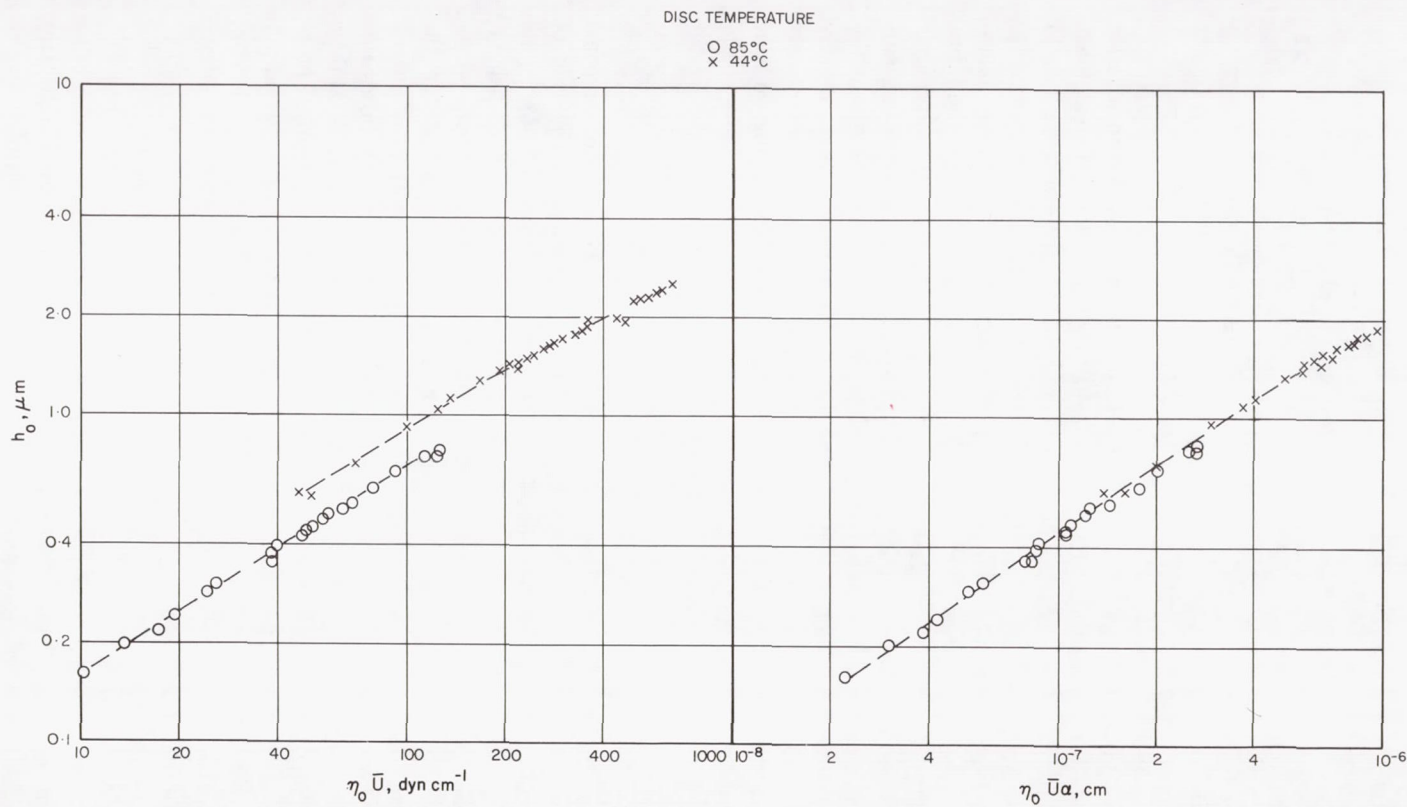


FIGURE 1.—Correlation of film thickness with (a)  $\eta_0 \bar{U}$  and (b)  $\eta_0 \bar{U} \alpha$ .

TABLE 3.—*Properties of Oil D at Two Different Temperatures*

Temperature, °C	Viscosity, $\eta_0$ , poise, at atmospheric pressure	Pressure coefficient of viscosity, $\alpha$ , $\text{dyn}^{-1} \text{cm}^2$ , over range 0 to 5 000 lbf in. <sup>-2</sup>
44	1.12	$2.98 \times 10^{-9}$
85	0.151	$2.17 \times 10^{-9}$

It is, perhaps, one of the disappointments of elastohydrodynamic theory that it has failed to provide a criterion for the design against scoring failure of heavily loaded, lubricated contacts. However, it would appear that in all cases in which lubricated Hertzian contacts operate under high loads with sliding, the mechanism of lubrication is elastohydrodynamic and that successful operation is only possible if the lubricant film remains intact. Nevertheless, in almost all elastohydrodynamic lubrication situations, some metallic contact between the opposing surfaces is observed. It is the nature and extent of these contacts that determine eventual failure. Thus the conditions for scoring failure are only loosely related to hydrodynamic factors, but are much more closely related to metallurgical and chemical phenomena. This being so, one cannot expect close correlation between, for example, the pressure coefficient of viscosity of the lubricant and the load at failure.

#### LECTURER'S CLOSURE

Prof. Klaus and Mr. Dyson again stress the importance of physical properties, and they urge a more thorough evaluation of the role of these properties in lubrication. I can only reiterate that I know of no case where one oil was selected over another on the basis of density, pressure-viscosity coefficient, wettability, specific heat, thermal conductivity, gas solubility, viscoelasticity, or volatility (except for the obvious case where a substantial portion of the lubricant evaporates). On the other hand, it is an everyday and routine occurrence for lubricants to be specified on the basis of oxidation resistance, thermal stability, detergency, flammability, and corrosion—all of which are chemical properties. It might be interesting to see how long it would take to compile a list of a dozen authenticated cases where one oil was chosen in preference to another on the basis of any physical property besides viscosity, evaporation, and freezing point.

It is necessary here to distinguish between the concept and the physical property. Heat transfer is obviously important in a bearing, but this does not necessarily mean that the thermal conductivity of the oil is crucial. Wetting is important (at least it was in one case), but the wetting properties of the oil apparently are not. At any rate, the problem in this

particular case was not solved by using a lubricant with better wettability (although it would have been easy to find a lubricant that could wet the surface) but rather by preventing contamination of the surface in the first place.

I can understand the appeal of physical properties; they are amenable to mathematical treatment whereas chemical properties generally are not. However, to concentrate on physical properties makes us focus on the wrong target; we try to find physical solutions to chemical problems. For example, a great deal of money is currently being spent to measure the viscosity of oils at extremely high pressures before there are any data to show that the pressure-viscosity characteristics are important even directionally.

Mr. Dyson has presented some new data that apply to this last point. However, in my opinion this experiment lacks being the critical experiment on two counts. First, in this test an oil was run at two different temperatures at which the viscosity differed by a factor of seven and the pressure-viscosity coefficient,  $\alpha$ , by a factor of 1.37. The tests were also run at different speeds to compensate for the difference in viscosity. Thus there were four variables—temperature, speed, viscosity, and  $\alpha$ . I would prefer to have the tests run at the same speed and at the same temperature, choosing oils having the same viscosity as the test temperature but different  $\alpha$ 's. I would also suggest three oils—a polyphenylether, a naphthenic white oil, and a long-chain synthetic ester. In this way,  $\alpha$  is the only variable, and its effect can be measured directly. We do not have to rely on the validity of the EHD equation nor make any assumptions about the true temperature in the area of contact.

The second limitation of the experiment is that it still measures film thickness. And as I pointed out, film thickness is a means to an end and not the end itself. I would prefer the experiment be carried to the point where there is some sign of distress—an increase in friction or some evidence of metal-to-metal contact. In such a test I would put my money on the synthetic ester, even though it has the lowest  $\alpha$  of the three.

Perhaps the tests could compare a naphthenic white oil with the same oil containing a few ppm of an acid phosphate additive. This additive will have no effect on viscosity or  $\alpha$  (the only oil properties in the EHD equation). But, as Prof. Klaus will aver, it will have a profound effect on the friction and wear characteristics.

Dr. Naylor's comments have summarized, better than I, the current status of elastohydrodynamic theory. The theory is fairly successful in predicting the thickness of oil films that are intact, but it cannot predict when the film will fail. The theory is philosophically satisfying by explaining how heavy loads can be carried without causing wear, but it does not help in selecting which oil will best lubricate a bearing.

The only other physical property worth mentioning is foaming, as



mentioned by Prof. Klaus (although there might be some argument whether this is truly a physical property or whether it is more related to the surface chemistry of the oil). However, I would not equate (as Prof. Klaus seems to have done) foaming and gas solubility. Foaming can be mitigated by a few ppm of silicone which does not lower gas solubility and may increase it.

Finally Mr. Dyson asks for clarification of Borsoff's experiment on the effect of speed on the load-carrying capacity of gears. The experimental data showed that load-carrying capacity first decreased with speed and then increased. Borsoff questioned the increase and suggested viscoelastic effects. We now know that viscoelastic oils give films that are thinner, not thicker. Therefore, his suggestion is not correct—which is the major point I was trying to make. I then went on to suggest that Borsoff may have questioned the wrong part of the curve. After all, from simple hydrodynamic theory, we would expect higher speeds to give thicker films and therefore better load-carrying capacity. It would appear harder to explain what happens at lower speeds where the load-carrying capacity decreased with speed. It seems to me that the likely answer is that the increase in speed caused an increase in temperature, and this in turn made the EHD film too thin to prevent metal-to-metal contact and scuffing. At high speeds, the EHD film was complete and did not allow the surfaces to touch. I do not think Mr. Dyson would disagree with this analysis.

#### REFERENCES

1. METRO, S. J.; HOFFMAN, E. M.; AND MATUSZAK, A. H.: Fatty Acid Esters of Polyhydric Alcohols. U.S. Patent 3 282 971, November 1966.
2. GUNDERSON, R. C.; AND HART, A. W.: Synthetic Lubricants. Reinhold Publishing Corp., 1962.
3. BOWERS, R. C.; AND MURPHY, C. M.: Status of Research on Lubricants, Friction and Wear. Naval Res. Lab. Rept. 6466, Jan. 1967.
4. BRAITHWAITE, E. R.: Solid Lubricants and Surfaces. The Macmillan Co., 1964.
5. BENEDICT, G. H.: Gears. Standard Handbook of Lubrication Engineering, J. J. O'Connor, ed., McGraw-Hill Book Co., Inc., 1968.
6. GODFREY, D.: Boundary Lubrication. Int. Symp. on Lubrication and Wear, D. Muster and B. Sternlicht, eds., McCutcheon, 1965.
7. LIPSON, C.: Wear Considerations in Design. Prentice-Hall, Inc., 1967.
8. ARCHARD, J. F.; AND HIRST, W.: The Wear of Metals Under Unlubricated Conditions. Proc. Roy. Soc. (London), vol. A236, 1956, p. 397.
9. HIRST, W.: The Mechanical Wear of Metals. Brit. J. Appl. Phys., vol. 9, 1958, p. 125.
10. LANCASTER, J. K.: The Formation of Surface Films and the Transition Between Mild and Severe Wear. Proc. Roy. Soc. (London), vol. A273, 1963, p. 466.
11. BURWELL, J. T., JR.: Survey of Possible Wear Mechanisms. Wear, vol. 1, 1958, p. 119.
12. LEVINE, O.; AND ZISMAN, W. A.: Physical Properties of Monolayers Adsorbed at the Solid-Air Interface I. Friction and Wettability of Aliphatic Polar Compounds and the Effect of Halogenation. J. Phys. Chem., vol. 61, 1957, p. 1063.

13. LANCASTER, P. R.; AND ROWE, G. W.: A Comparison of Boundary Lubricants Under Light and Heavy Loads. *Wear*, vol. 2, 1959, p. 428.
14. APPELDOORN, J. K.; TAO, F. F.; AND GOLDMAN, I. B.: Corrosive Wear By Atmospheric Oxygen and Moisture. *ASLE Trans.*, vol. 12, 1969, p. 140.
15. TAMAI, Y.: The Interactions of Wear Debris with Fatty Additives in Lubrication. *Wear*, vol. 2, 1958–59, p. 228.
16. ZISMAN, W. A.: Historical Review, Lubricants and Lubrication. *Synthetic Lubricants*, R. C. Gunderson, ed., Reinhold Publishing Corp., 1962.
17. ASKWITH, T. C.; CAMERON, A.; AND CROUCH, R. F.: Chain Length of Additives in Relation to Lubricants in Thin Film and Boundary Lubrication. *Proc. Roy. Soc. (London)*, vol. A291, 1966, p. 500.
18. FEIN, R. S.: Effects of Lubricant on Transition Temperatures. *ASLE Trans.*, vol. 8, 1965, p. 59.
19. LEVINE, O.; AND ZISMAN, W. A.: Physical Properties of Monolayers Adsorbed at the Solid-Air Interface II. Mechanical Durability of Aliphatic Polar Compounds and Effect of Halogenation. *J. Phys. Chem.*, vol. 61, 1957, p. 1188.
20. APPELDOORN, J. K.; AND TAO, F. F.: The Lubricity Characteristics of Heavy Aromatics. *Wear*, vol. 12, 1968, p. 117.
21. KLAUS, E. E.; AND BIEBER, H. E.: Effect of Some Physical and Chemical Properties of Lubricants on Boundary Lubrication. *ASLE Trans.*, vol. 7, 1964, p. 1.
22. FEIN, R. S.; AND KREUZ, K. L.: Chemistry of Boundary Lubrication of Steel by Hydrocarbons. *ASLE Trans.*, vol. 8, 1965, p. 29.
23. ANON.: ASTM Method D-2390: Microscopic Sizing and Counting Particles from Aerospace Fluids on Membrane Filters. *ASTM Standards*, ASTM, Philadelphia, Pa.
24. GODFREY, D.: Investigation of Fretting Corrosion by Microscopic Observation. *NACA TN 2039*, February 1950.
25. LANDEN, E. W.: Slow Speed Wear of Steel Surfaces Lubricated by Thin Oil Film. *ASLE Trans.*, vol. 11, 1968, p. 6.
26. BORSOFF, V. N.: Gear Lubrication—A Descriptive Analysis. *Lub. Engrg.*, vol. 18, 1962, p. 266.
27. BLOK, H.: General Discussion on Lubrication. *Inst. Mech. Engrs.*, vol. 2, 1937, p. 14.
28. ARCHARD, J. F.: The Temperature of Rubbing Surfaces. *Wear*, vol. 2, 1959, p. 438.
29. LEACH, E. F.; AND KELLEY, B. W.: Temperature—The Key to Lubricant Capacity. *ASLE Trans.*, vol. 8, 1965, p. 271.
30. ANON.: ASTM Method D-2422: Viscosity System for Industrial Fluid Lubricants. *ASTM Standards*, ASTM, Philadelphia, Pa.
31. BOWDEN, F. P.; AND TABOR, D.: The Friction and Lubrication of Solids, Part II. Clarendon Press (Oxford), 1964, p. 367.
32. MANTEUFFEL, A. A.; AND WOLFRAM, G.: A Method for Studying the Effect of Extreme Pressure Additives on Rubbing Metal Surfaces. *ASLE Trans.*, vol. 3, 1960, p. 157.
33. PILPEL, N.: Extreme Pressure Lubricants. *Research (London)*, vol. 12, 1959, p. 141.
34. BONER, C. J.: Gear and Transmission Lubricants. Reinhold Publishing Corp., 1964.
35. MOUGEY, H. C.; AND ALMEN, J. D.: Extreme Pressure Lubricants. *Proc. API*, vol. 11, III (12), 1931, p. 76.
36. HENDRICK, E. E.: U.S. Patent 90 100, May 18, 1869.
37. GODFREY, D.: Boundary Lubrication. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, p. 335.



38. ROWE, C. N.; AND DICKERT, J. J., JR.: The Relation of Antiwear Function to Thermal Stability and Structure for Metal O, O, Dialkylphosphorodithioates. ASLE Trans., vol. 10, 1967, p. 85.
39. PRUTTON, C. F.; TURNBULL, D.; AND DLOUHY, G.: Mechanism of Action of Organic Chlorine and Sulfur Compounds in Extreme Pressure Lubrication. J. Inst. Petr., vol. 32, 1946, p. 90.
40. SAKURAI, T.; AND SATO, K.: Study of Corrosivity and Correlation Between Chemical Reactivity and Load Carrying Capacity of Oils Containing Extreme Pressure Agents. ASLE Trans., vol. 9, 1966, p. 77.
41. BISSON, E. E.; SWIKERT, M. A.; AND JOHNSON, R. L.: Effect of Chemical Reactivity of Lubricant Additives on Friction and Surface Welding at High Sliding Velocities. NACA TN 2144, Aug. 1950.
42. CAMPBELL, R. B.: Sulphur as an Extreme Pressure Lubricant. Sci. Lub., vol. 9, 1957, p. 109.
43. DORINSON, A.; AND BROMAN, V. E.: Extreme Pressure Lubrication & Wear. The Chemical Reactivity and the Extreme Pressure Action of Two Aliphatic Disulfides. ASLE Trans., vol. 5, 1962, p. 75.
44. FUREY, M. J.; AND KUNC, J. F., JR.: A Radiotracer Approach to the Study of Engine Valve Train Lubrication. Petr. Div. Preprints, ACS meeting, Miami, Fla., April 7-12, 1957.
45. LOESER, E. H.; WIKQUIST, R. C.; AND TWISS, S. B.: Cam & Tappet Lubrication. III. Radioactive Study of Phosphorus in the EP Film. ASLE Trans., vol. 1, 1958, p. 329.
46. ROUNDS, F. G.: Some Effects of Nonhydrocarbon Base Oil Constituents on the Friction and Surface Coating Formation Obtained with Three Additives. ASLE Trans., vol. 11, 1968, p. 19.
47. ALLUM, K. G.; AND FORBES, E. S.: The Load Carrying Mechanism of Organic Sulfur Compounds—Application of Electron Probe Microanalysis. ASLE Trans., vol. 11, 1968, p. 162.
48. DAVEY, W.; AND EDWARDS, E. D.: The Extreme Pressure Lubricating Properties of Some Sulphides and Disulphides, in Mineral Oil, as Assessed by the Four Ball Machine. Wear, vol. 1, 1957-58, p. 291.
49. BARCROFT, F. T.: A Technique for Investigating Reactions Between EP Additives and Metal Surfaces at High Temperatures. Wear, vol. 3, 1960, p. 440.
50. BEECK, O.; GIVENS, J. W.; AND WILLIAMS, E. C.: On the Mechanism of Boundary Lubrication. II. Wear Prevention by Addition Agents. Proc. Roy. Soc. (London), vol. 177A, 1940, p. 103.
51. CAMPBELL, W. E.: Studies On Boundary Lubrication. Trans. ASME, vol. 61, 1939, p. 633.
52. GODFREY, D.: The Lubrication Mechanism of Tricresyl Phosphate on Steel. ASLE Trans., vol. 8, 1965, p. 1.
53. GODFREY, D.: Chemical Changes in Steel Surfaces During Extreme Pressure Lubrication. ASLE Trans., vol. 5, 1962, p. 5.
54. STERNLICHT, B.; LEWIS, P.; AND FLYNN, P.: Theory of Lubrication and Failure of Rolling Contacts. Trans. ASME, J. Basic Engrg., vol. 83, 1961, p. 213.
55. DAWSON, P. H.: The Effect of Metallic Contact On The Pitting Of Lubricated Rolling Surfaces. J. Mech. Engrg. Sci., vol. 4, 1962, p. 16.
56. TALLIAN, T. E.; MCCOOL, J. I.; AND SIBLEY, L. B.: Partial Elastohydrodynamic Lubrication in Rolling Contact. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965-66, p. 169.
57. HARRIS, T. A.: Load-Life Characteristics of Rolling Bearings. Standard Handbook of Lubrication Engineering, J. J. O'Connor, ed., McGraw-Hill Book Co., Inc., 1968, pp. 6-14.



58. BARWELL, F. T.; AND SCOTT, D.: Effect of Lubricant on Pitting Failure of Ball Bearings. *Engineering*, vol. 182, 1956, p. 9.
59. OTTERBEIN, M. E.: Effect of Aircraft Gas Turbine Oils on Roller Bearing Fatigue Life. *ASLE Trans.*, vol. 1, 1958, p. 33.
60. APPELDOORN, J. K.; AND ROYLE, R. C.: Lubricant Fatigue Testing with Ceramic Balls. *Lub. Engrg.*, vol. 21, 1965, p. 45.
61. ROUNDS, F. G.: Effects of Base Oil Viscosity and Type on Bearing Ball Fatigue. *ASLE Trans.*, vol. 5, 1962, p. 172.
62. ROUNDS, F. G.: Some Effects of Additives on Rolling Contact Fatigue. *ASLE Trans.*, vol. 10, 1967, p. 243.
63. KOVED, I.: A Comparison of Fatigue Test Techniques for Gas Turbine Oils SAE paper 680322. Presented at Air Transportation Meeting, New York, New York, April 29–May 2, 1968.
64. GRUNBERG, L.; AND SCOTT, D.: The Acceleration of Pitting Failure by Water in the Lubricant. *J. Inst. Petr.*, vol. 44, 1958, p. 406.
65. SCHATZBERG, P.; AND FELSEN, I. M.: Effects of Water and Oxygen During Rolling Contact Lubrication. *Wear*, vol. 12, 1968, p. 331.
66. ANON.: ASME Pressure-Viscosity Report, ASME, New York, 1953.
67. ANON.: API Res. Proj. 42: Properties of Hydrocarbons of High Molecular Weight. API, New York, 1967.
68. KANNEL, J.; BELL, J. C.; WALOWIT, J. A.; AND ALLEN, C. M.: A Study of the Influence of Lubricants on High-Speed Rolling-Contact Bearing Performance, Part VIII. ASD-TDR-61-643, June, 1968.
69. ANDERSON, W. J.; AND CARTER, T. L.: Effect of Lubricant Viscosity and Type on Ball Fatigue Life. *ASLE Trans.*, vol. 1, 1958, p. 266.
70. ZISMAN, W. A.: Relation of Equilibrium Contact Angle to Liquid and Solid Constitution. Contact Angle, Wettability, and Adhesion, *Adv. in Chem. Series 43*, ACS, Washington, 1964.
71. FOX, H. W.; HARE, E. F.; AND ZISMAN, W. A.: Wetting Properties of Organic Liquids on High Energy Surfaces. *J. Phys. Chem.*, vol. 59, 1955, p. 1097.
72. BENNETT, M. K.; AND ZISMAN, W. A.: Prevention of Liquid Spreading or Creeping. Contact Angle, Wettability, and Adhesion, *Adv. in Chem. Series 43*, ACS, Washington, 1964.
73. FREEMAN, A. P.; ALLEN, S.; AND SINGER, H. B.: Ball Bearing Surface Chemistry. Report E-2288, MIT Instrumentation Lab, Cambridge, Mass.
74. GASKINS, F. H.; AND PHILIPPOFF, W.: Instrumentation for Rheological Investigation of Viscoelastic Materials. *Ind. Eng. Chem.*, vol. 51, 1959, p. 871.
75. PHILIPPOFF, W.; AND STRATTON, R. A.: Correlation of the Weissenberg Rheogoniometer with Other Methods. *Trans. Soc. Rheol.*, vol. 10, 1966, p. 467.
76. APPELDOORN, J. K.; AND PHILIPPOFF, W.: Rheology of Polymer Solutions in High Viscosity Solvents. *Preprints of Petr. Div., ACS meeting, Atlantic City, N. J., Sept. 1962*.
77. TANNER, R. I.: Response of Viscoelastic Fluids in Dynamically Loaded Bearings. *Trans. ASME, J. Lub. Tech.*, vol. 90F, 1968, p. 555.
78. HOROWITZ, H. H.: Predicting Effects of Temperature and Shear Rate on Viscosity of V.I. Improved Lubricants. *Ind. Eng. Chem.*, vol. 50, 1958, p. 1089.
79. ANON.: Union Carbide Corp. Design Sheet No. 3: Apparent Viscosity vs. Shear Rate. July 1, 1959.
80. BARLOW, A. J.; AND LAMB, J.: Viscoelastic Behavior of Lubricating Oils Under Cyclic Shearing Stress. *Proc. Roy. Soc. (London)*, vol. A253, 1959, p. 52.
81. WEISSENBERG, K.: Seminar On Continuum Mechanics. Columbia Univ., ONR Project NR 064-446, Tech. Report No. 18, 1963.
82. APPELDOORN, J. K.: Physical Properties of Lubricants. *Boundary Lubrication*,

- An Appraisal of World Literature, F. F. Ling, E. E. Klaus, and R. S. Fein, eds., ASME, New York, ch. 8., 1969.
83. TAO, F. F.; AND PHILIPPOFF, W.: Hydrodynamic Behavior of Viscoelastic Liquids in a Simulated Journal Bearing. ASLE Trans., vol. 10, 1967, p. 302.
  84. CAMERON, A.: Surface Failure in Gears. J. Inst. Petr., vol. 40, 1954, p. 191.
  85. BURTON, R. A.: Analytical Investigation of Viscoelastic Effects in the Lubrication of Rolling Contact. ASLE Trans., vol. 3, 1960, p. 1.
  86. MILNE, A. A.: A Theory of Rheodynamic Lubrication for a Maxwell Liquid. Proc. of Conf. of Lubr. and Wear, Inst. Mech. Engrs., London, 1957.
  87. MOW, V. C.: Effects of Viscoelastic Lubricant on Squeeze Film Lubrication Between Impinging Spheres. Trans. ASME, J. Lub. Tech., vol. 90F, 1968, p. 113.
  88. BORSOFF, V. N.: Mechanism of Gear Lubrication. Trans ASME, J. Basic Engrg., vol. 81D, 1959, p. 79.
  89. FUREY, M. J.; AND APPELDOORN, J. K.; The Effect of Lubricant Viscosity on Metallic Contact and Friction in a Sliding System. ASLE Trans., vol. 5, 1962, p. 149.
  90. DYSON, A.; NAYLOR, H.; AND WILSON, A. R.: The Measurement of Oil Film Thickness in Elastohydrodynamic Contacts. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965-66, p. 119.

# Adsorption and Surface Energetics

**A. J. HALTNER**

**General Electric Company  
Philadelphia, Pennsylvania**

The surface properties of solids are critical factors affecting the course of friction and wear in concentrated contacts. The physical and chemical states of the surface affect such processes as surface deformation and interaction with the environment. Knowledge concerning details of ideal surfaces is rapidly growing as a consequence of the introduction of sophisticated techniques for surface examination such as low energy electron diffraction, field ion microscopy, electron microprobe analysis, and scanning electron microscopy. It is now possible to study in detail the atomistics of important surface processes. The thermodynamics of solid surfaces are complex because of the physical and chemical inhomogeneities present at real surfaces. However, surface energy remains a useful parameter for interpreting and predicting interactions that occur at interfaces such as wetting and spreading. Surface energies for solids range from around 20 ergs/cm<sup>2</sup> (low energy surfaces) to more than 2000 ergs/cm<sup>2</sup> (high energy metal surfaces). The range of surface forces depends on the various types of force fields that may be involved, and these can vary widely from one type of surface to another. The application of equilibrium thermodynamic analysis to surfaces involved in sliding has considerable validity. However, limitations on such analyses may be imposed by kinetic restrictions inherent in the thermodynamically allowed processes that may be occurring in the interface region.

**A** LOADED JOURNAL OR THRUST BEARING operating within its design limitations distributes the applied load continuously (though not necessarily uniformly) over a rather large macroscopic area. In ball and roller bearings, gears, and a large variety of sliding systems, thick oil films do not separate the relatively moving surfaces under high load. The geometric planes defining average surface positions approach to distances comparable to average asperity heights. As a consequence, the real load bearing area becomes more restricted in extent and can consist of an assemblage of discrete contacts within the contact zone. The actual extent of the load bearing area begins to be dictated by the mechanical properties of the contacting surfaces.

In the friction and wear in such concentrated contact systems, par-



ticularly where oily lubricants are introduced, it is common to speak of boundary lubrication (refs. 1 and 2). A precise definition of this term is a matter of some debate (refs. 3 and 4). In the classical picture of boundary lubrication, it is generally assumed that extensive penetration of the liquid lubricant film by surface asperities has occurred to permit solid surfaces to come into contact. It is then the assumed role of the boundary lubricant to modify the properties of the contacting surfaces such that the course of friction and wear is mitigated. Since a chemical role is assigned to the lubricant, it is apparent that mechanisms of boundary lubrication should be complex, involving interactions with contaminated surfaces as well as with nascent surfaces formed by wear processes.

Boundary lubrication is often considered not to encompass dry sliding. But when sliding occurs in air or reactive atmospheres, there is evidence of interaction of gas molecules with the solid surfaces to limit the extent of metal junctions and limit friction and wear under conditions of mild wear. This is accomplished presumably by mechanisms similar to those encountered in sliding in the presence of liquid lubricants.

Recently there has been a growing tendency to consider that the boundary lubrication regime does not merge directly into the hydrodynamic regime (ref. 5). It has been suggested that, under conditions where it had previously been supposed that solid contacts were dominating friction and wear, there is the possibility that the surfaces are still separated by a thin film of fluid lubricant (refs. 6 and 7). In an extension of the theory to micro-asperity contacts (ref. 8), enormous pressures limited by asperity deformation are anticipated between opposing asperity surfaces resulting in significant increases in fluid viscosity. The theory of elastohydrodynamic lubrication is developed from hydrodynamic theory by including terms in the Reynold's equation to describe the surface deformation and the pressure viscosity behavior of the oil.

Elastohydrodynamic theory would appear to have a distinct advantage over boundary lubrication theory because of the quantitative nature of the former's impressive mathematical formalism. Boundary lubrication has grown to include a descriptive, qualitative body of knowledge much like the science of chemistry before the introduction of chemical thermodynamics and wave mechanics. But there are also difficulties with EHD theory. This theory is still in its infancy, and there are formidable difficulties present in attempts to solve the differential equations constructed to describe real lubrication situations. Solutions are then based on assumptions that simplify the boundary conditions and adapt idealized representations of surface contours (ref. 9).

Whether the boundary lubrication or the EHD picture of a concentrated contact is accepted, and it is more than likely that each has validity depending on specific conditions, it is obvious that surface properties play central roles in friction and wear mechanisms. The

deformation processes of metals are strongly determined by environment, the presence of surface films, and the presence of surfactants (refs. 10 and 11). An interesting conclusion of reasoning based on EHD theory is that in contrast to the pure Hertzian contact case the maximum shear stress for EHD conditions occurs about a factor of 10 closer to the surface (ref. 12). Therefore, in the case of the deformation of micro-asperities having relatively small radii of curvature, the maximum shear stress should be in the surface region, and deformation and wear processes should be dependent on properties of the surface region.

A difficulty in the extensive boundary lubrication literature is the apparent discrepancies in the data and the lack of reproducibility from one investigator to another. It is obvious in classical boundary lubrication theory that there should be a close relationship between friction and wear in this lubrication regime and the properties of the solid surfaces in contact as well as their chemical interactions with the lubricant. This picture has been discussed at length by many authors (refs. 3, 13, and 14). Now it should be apparent that measurements of a surface chemical nature are most easily accomplished in the laboratory. They require no elaborate facilities or apparatus. Thus investigators interested in boundary lubrication commonly make measurements of wetting, spreading, and contact angle. Perhaps it has not been sufficiently stressed that such experiments are extremely difficult to control. This is because the phenomena being observed depend directly on interface properties, and the interface is a very small fraction of the total system. Small quantities of unsuspected foreign agents from liquids, solids, or the surrounding atmosphere can concentrate at the interface, resulting in significant changes in interface properties. Boundary lubricated systems are extremely complex, and rarely is the system under such complete experimental control that it can be described adequately. Without such control the body of information in boundary lubrication remains uninterpretable, and a unified quantitative theory of boundary lubrication is not possible.

This brief introduction has emphasized the dominance of the surface properties of solids in determining the course of friction and wear when sliding conditions are outside the hydrodynamic regime. This is obviously the case in dry sliding, but it should also apply if a boundary lubrication or EHD regime or some transition region between them is involved. It is the purpose of the remainder of this essay to consider surface properties of solids in terms of lubrication problems. Emphasis will be placed on recent developments in surface science that have implications for the lubrication of concentrated contacts.

#### NATURE OF SOLID SURFACES

Surfaces of real solids are not as a rule homogeneous. One source of this inhomogeneity arises in the regularity with which atoms arrange



themselves in crystalline solids. In even the simplest of crystals the solid can be bounded by a variety of planes meeting at definite angles with each plane differing in the density of atoms present and the type of packing. Where a crystalline compound is considered, these planes may differ also in chemical composition. Graphite may serve as an extreme example where the basal plane is significantly different from all other possible planes in the crystal. In a given crystal, the properties of these various surfaces would be expected to vary widely. This has been verified by numerous experiments with clean crystal surfaces generated in ultra-high vacuum by cleavage or other methods.

Real single crystals are not perfect. They contain defects or dislocations that influence the properties of the various possible crystal surfaces. Most real solids are polycrystallines, and as a result one is usually dealing with an assemblage of crystal planes.

Real surfaces, even if originally clean, are rapidly contaminated (ref. 15). Even in high vacuum ( $10^{-6}$  torr), a chemisorbed layer will form within seconds. In ordinary ambient atmosphere, nascent surfaces immediately oxidize, and thick oxide layers form almost instantaneously. Only in relatively rare instances are well-characterized surfaces produced and studied (ref. 16). The recent development of ultrahigh vacuum techniques has made such studies more convenient, but only rarely has such care been applied to studies related to the lubrication of concentrated contacts.

Fundamental knowledge of surfaces and the processes that occur there has been advanced greatly by the development of a number of sensitive physical techniques. These techniques tell us how atoms in the surface are arranged, allow us to identify impurities present in less than monolayer quantities, allow a determination of the arrangement of impurity atoms as a function of crystal surface, and permit an assessment of the relative degrees of stability of foreign atoms on various crystalline planes. It is now possible to follow single atomic events on crystal surfaces, so many of the myriad processes that are possible at surfaces can be followed in detail under a variety of carefully controlled conditions. In principle, model systems that resemble complex boundary lubrication situations can now be unraveled given sufficient time and patience.

Measurement of the work function (ref. 17), which is a measure of the ease with which electrons can be removed from a metal surface, is a useful tool to characterize a metal surface since the work function can vary from crystal plane to crystal plane and is sensitive to the presence of contamination. For this reason, measurements of the work function are often used as a test of the surface cleanliness and homogeneity. Low energy electrons will penetrate into a metal to a depth of only two or three atomic layers. Consequently the resulting diffraction pattern from a single crystal surface will be a two-dimensional map of the surface.



Low energy electron diffraction (LEED) (ref. 18) confirms that the surfaces of metals have structures similar to the bulk, but are consistent with some expansion of the lattice in the surface region along a normal to the surface. For covalent solids considerable reconstitution of the surface region is noted (ref. 19). This presumably occurs to partially satisfy severed valence bonds otherwise left dangling at the surface.

The field emission microscope (ref. 20), which operates in ultrahigh vacuum, uses a high field to extract electrons from a fine tip of a refractory metal and accelerates them to a fluorescent screen. The resulting image is a work function map of the tip surface. The regions of the various crystalline planes exposed in the tip are readily seen; but since the resolution is limited to  $20\text{\AA}$ , individual atomic sites cannot be seen. In the field ion microscope (ref. 21), ionized helium is accelerated to the screen, and the resolution is improved to about  $2\text{\AA}$ . The resulting patterns reveal individual atomic sites, and it has been possible to follow individual adatoms on the tip surface. Because of the uniformity of low index planes in the tip, these do not produce detailed images. When hydrogen is used as the imaging gas, details of these planes are revealed (ref. 22). In addition, field evaporation of the tip can proceed at greatly reduced voltages so the technique can be applied to less refractory metals such as iron, nickel, cobalt, and carbon steel. These, of course, are materials of more immediate interest as bearing surfaces. An intriguing application of the microscope is to study surface deformation and adhesion of tip materials as a function of load (ref. 23). Plastic damage and the presence of dislocations are observed directly and in detail. By using controlled field evaporation, the subsurface damage can be studied layer by layer and its extent determined. In an instance where the load on a tungsten contact did not produce plastic damage and an electric current was passed through the contact, there was extensive surface disordering in the contact zone. Many of the factors presumed to be operating in concentrated lubricated contacts such as surface contamination can be carefully controlled by this method, and effects on deformation and transfer can be observed. The method is complex and interpretation of data is not straightforward, but the promise appears great.

A sensitive technique that is beginning to be applied to problems of friction and wear is the electron microprobe (ref. 24). In the instrument, surface atoms excited by electron bombardment emit characteristic x-rays that permit both qualitative and quantitative analysis of the elements in the surface. By scanning techniques a map of the distribution of the elements in the surface can be generated.

These techniques and others are dramatically advancing our knowledge of surfaces. Since lack of detailed information about surfaces is a limitation to the development of successful theories of friction and wear, this new knowledge should have beneficial impact in the field of lubrication.

## SURFACE ENERGETICS

All molecules attract each other to a greater or lesser degree. These universal attractive forces account for the cohesive strengths of liquids and solids and are also at the basis of the familiar fluid phenomena of capillarity such as the rise of liquids in fine bore tubes and the pressure differential across a curved surface. Work is required to create new surface against the cohesive forces, and the reversible work required to create one square centimeter of new surface is the surface free energy. It is designated by  $\gamma$  and for liquids is equal numerically to the surface tension. When only the surface area is increased, the total free energy of the system is changed by

$$dG = \gamma dA \quad (1)$$

where  $dG$  is the change in free energy and  $dA$  is the change in area. Therefore, the surface free energy,  $G^s$ , is given by

$$G^s = (\partial G / \partial A)_{T,P,n} \quad (2)$$

and the notation indicates that the partial derivative is to be taken at constant temperature, pressure, and composition. The heat adsorbed in the extension of the surface,  $dq$ , is given by  $TS^s dA$  where  $S^s$  is the surface entropy. Then

$$(\partial G^s / \partial T)_p = -S^s = (\partial \gamma / \partial T) \quad (3)$$

According to the second law of thermodynamics,

$$H^s = G^s + TS^s \quad (4)$$

where  $H^s$  is the enthalpy,  $H^s = E^s + PV^s$  and  $E^s = q^s + w$  with  $w$  including all energy terms other than heat. For surfaces one may equate  $H^s$  with  $E^s$ , the total surface energy, since  $V^s$  is practically zero. Then we find on substitution that

$$E^s = \gamma - T(\partial \gamma / \partial T) \quad (5)$$

Therefore, the total surface energy is always larger than the surface free energy by an entropy term that will be positive because surface tension decreases with temperature. A successful empirical relationship for variation of surface tension of liquids with temperature is (ref. 25)

$$\gamma = \gamma_o(1 - T_r)^{11/9} \quad (6)$$

$T_r$  is  $T/T_c$  with  $T_c$  the critical temperature, and  $\gamma_o$  is a constant.

The above equations adequately describe a pure liquid. When more than one component is present, the thermodynamic equations become more complex. The basic equation for the surface tension becomes

$$d\gamma = -S^s dT - \sum \Gamma_i d\mu_i \quad (7)$$



where  $\Gamma_i$  and  $\mu_i$  are the surface concentration and chemical potential of the  $i$ 'th component in the system. For a solvent single solute combination, the Gibb's adsorption equation gives

$$\Gamma_2 = - \frac{c_2}{RT} \frac{\partial \gamma}{\partial c_2} \quad (8)$$

If the excess surface concentration is positive, then surface energy will decrease as  $c_2$  increases. But if the excess surface concentration is negative, the opposite will occur.

Molecules in liquid surfaces are mobile, and these surfaces will equilibrate readily to produce homogeneous surfaces in which the transition across the surface region occurs in a dimension of the order of a molecular diameter. Atoms in solid surfaces are not very mobile, and surfaces of real solids often display atomic arrangements that do not constitute the equilibrium (lowest energy) state. The surface free energy is a definite thermodynamic quantity as defined above, but it is often difficult to get at experimentally. The total surface energy of real solids will contain contributions from such sources as surface imperfections and non-equilibrium arrangements of surface atoms (strain) for the relief of which there is no practical kinetic mechanism. These factors introduce uncertainties into experimental methods that seek to get at surface free energy by measuring heat capacity or heat of solution of small particles relative to larger ones (ref. 26). Early attempts, both theoretical and experimental, to estimate the surface energies of solids are summarized by Partington (ref. 27). The surface free energy of solids near the melting point would not be expected to differ greatly from the surface tensions of the corresponding liquids at the melting points. Bondi (ref. 28) estimates this correction in surface free energy on solidification as

$$\Delta \gamma = \Delta h_f / A_w \quad (9)$$

where  $\Delta h_f$  is the heat of fusion and  $A_w$  is the area per atom in the interface. Typically for metals this correction appears to be of the order of 20 percent. A number of methods have been devised for measuring the surface free energy of metals. One method employs the measure of creep rate in a cylindrical single crystal metal wire near the melting point (ref. 29). That load at which creep is zero just balances the surface tension forces, and the surface free energy can be calculated. In another equilibrium technique, the tendency of a field emitter tip to shrink is balanced by applied electrostatic force (ref. 30). The surface free energy can also be deduced from the damping of sinusoidal oscillations in a single crystal surface. The technique has been applied to nickel surfaces (ref. 31). A dynamic method involves the following of a cleavage crack propagation (ref. 32).

For comparison, literature values for the surface free energies of repre-



sentative liquids and solids are gathered in table 1. The reported values cover a wide range. Inorganic and organic liquids have very low surface energy. Liquid metals have surface free energies that are almost two orders of magnitude higher. Solid metals have quite high energy surfaces, and as is expected these values vary with the crystal plane. When metal surfaces are reacted with oxygen or sulfur, the surface free energy is greatly reduced. Solid hydrocarbons or polymeric materials have quite low surface energies.

It is not correct to consider even a perfect single crystal surface as an assemblage of sites possessing equal energy. This may be seen quite easily for a surface made up of a simple square lattice (ref. 56). An atom approaching this surface can meet with a number of different energy conditions. Over a center of a square an atom interacts directly with the largest number of surface atoms, and this is a site having high interaction energy. An atom at a midpoint along an edge of the square interacts less strongly, and an atom directly over a surface atom has the lowest interaction energy of all. There are other more serious contributions to surface energy heterogeneity of real surfaces. These include surface imperfections such as dislocations, the presence of impurity atoms at the surface, and the existence of steps or terraces in the surface. Of course, another source is the distinct difference in energies from one crystal surface plane to another (table 1). These differences in the energies of surface sites lead to variations in the heat of adsorption as the degree of surface coverage varies. Fresh wear surfaces adsorb gases strongly as is demonstrated by the strong adsorption of wear debris from graphite for hydrogen (ref. 57). Quantitative information about the differences in adsorption site energies is furnished by flash desorption experiments (ref. 58). In these experiments gas atoms are adsorbed on fine filaments (such as tungsten), and then the rate of gas evolution is observed as the surface is rapidly heated (flashed). The gas does not all come off at a single temperature, but is evolved in discrete peaks with the weakly bound atoms being evolved at lower temperatures.

Various types of interaction are possible at solid surfaces and these give rise to the two familiar types of adsorption—physical and chemical. Strong energy reductions result from metallic bond interactions, electron exchange (homopolar bond formation), and ion-ion interactions. These interactions leading to chemical adsorption have steep energy minima with the force fields dropping off rapidly at separations of a few atomic radii. Less specific are the forces leading to hydrogen bond formation, dipole-dipole interactions, and induced dipole-induced dipole interactions. These forces lead to the familiar less energetic phenomena of physical adsorption. The induced dipole forces are often referred to as the London dispersion forces. They are quite generally present at surfaces and can be the determining factor in interactions across interfaces.

TABLE 1.—*Surface Energies*

Material	Surface energy, ergs/cm <sup>2</sup>	T, °C	Method	Ref.
<u>Liquids</u>				
1. Argon	13.2	-188	-----	33
2. n-Hexane	18.43	20	-----	33
3. Benzene	28.85	20	-----	33
4. Glycerol	63.4	20	-----	33
5. Water	72.75	20	-----	33
6. Hydrazine	91.5	25	-----	33
7. Selenium	92.4	217	-----	33
8. Silicone fluid SF-96-1000cs. $M_n = 6350$	24	20.6	Pendant drop	34
9. Hexamethyl disiloxane	24	15.3	Pendant drop	34
10. Polyisobutylene $M_n = 410$	24	28.4	Pendant drop	35
11. Polyisobutylene $M_n = 2840$	24	33.4	Pendant drop	35
12. B <sub>2</sub> O <sub>3</sub>	79.5	900	Ring pull	36
13. PbO	132	900	Ring pull	36
14. Na <sub>2</sub> SiO <sub>3</sub>	310	1000	-----	28
15. Al <sub>2</sub> O <sub>3</sub>	580	2050	-----	28
16. FeO	585	1420	-----	28
17. KI	75	737	Maximum bubble pressure (N <sub>2</sub> )	37
18. NaI	86	706	Maximum bubble pressure (N <sub>2</sub> )	37
19. CsI	73	654	Maximum bubble pressure (N <sub>2</sub> )	37
20. KCl	96	800	Maximum bubble pressure (N <sub>2</sub> )	37
21. CsF	104	722	Maximum bubble pressure (N <sub>2</sub> )	37
22. NaCl	114	803	Maximum bubble pressure (N <sub>2</sub> )	37
23. LiCl	138	604	Maximum bubble pressure (N <sub>2</sub> )	37
24. LiF	750	868	Maximum bubble pressure (N <sub>2</sub> )	37
25. Cs <sub>2</sub> SO <sub>4</sub>	111	1036	Maximum bubble pressure (N <sub>2</sub> )	37
26. Na <sub>2</sub> SO <sub>4</sub>	195	900	Maximum bubble pressure (N <sub>2</sub> )	37
27. Li <sub>2</sub> SO <sub>4</sub>	224	860	Maximum bubble pressure (N <sub>2</sub> )	37
28. Potassium	119	64	-----	38
29. Bismuth	376	300	-----	38

TABLE 1.—*Surface Energies—Continued*

Material	Surface energy, ergs/cm <sup>2</sup>	T, °C	Method	Ref.
30. Antimony	383	635	-----	38
31. Lead	442	350	-----	38
32. Mercury	476.1	20	-----	39
33. Cadmium	608	370	-----	38
34. Gallium	735	30	-----	28
35. Aluminum	900	700	-----	28
36. Silver	923	995	Maximum bubble pressure (H <sub>2</sub> )	40
37. Copper	1120	1140	Maximum bubble pressure (H <sub>2</sub> )	40
38. Gold	1128	1120	Maximum bubble pressure (H <sub>2</sub> )	40
39. Steels	950 to 1220		-----	41
40. Iron	1700	1530	Extrapolation to zero alloy concentration	28
<u>Solids</u>				
41. Polyhexafluoropropylene	19	25	Contact angle measurements $\cos \theta = 2\Phi \sqrt{\gamma_{so}/\gamma_L} - 1$	42
42. PTFE	24	25	Contact angle measurements $\cos \theta = 2\Phi \sqrt{\gamma_{so}/\gamma_L} - 1$	42
43. Poly(trifluoro monochloro) ethylene	38	25	Contact angle measurements $\cos \theta = 2\Phi \sqrt{\gamma_{so}/\gamma_L} - 1$	42
44. PTFE	34	25	Contact angle measurements $\gamma_3 = \pi_{32} + \gamma_{23} + \phi_{23}$	43
45. Paraffin	45	25	Contact angle measurements $\gamma_3 = \pi_{32} + \gamma_{23} + \phi_{23}$	43
46. Polyethylene	55	25	Contact angle measurements $\gamma_3 = \pi_{32} + \gamma_{23} + \phi_{23}$	43
47. Cellulose	200	20	-----	44
48. CaCO <sub>3</sub> (1010)	230	-196	Crystal cleavage (N <sub>2</sub> ) $\gamma = 6FL^2/E\mu^{2/3}$	32
49. BaF <sub>2</sub> (111)	280	-196	Crystal cleavage (N <sub>2</sub> )	32
50. LiF (100)	340	-196	Crystal cleavage (N <sub>2</sub> )	32
51. CaF <sub>2</sub> (111)	450	-196	Crystal cleavage (N <sub>2</sub> )	32
52. MgO (100)	1200	-196	Crystal cleavage (N <sub>2</sub> )	32
53. MgO	1040	25	Heat capacity measurements	26



TABLE 1.—*Surface Energies—Concluded*

Material	Surface energy, ergs/cm <sup>2</sup>	T, °C	Method	Ref.
54. Si (111)	1240	−196	Crystal cleavage (N <sub>2</sub> )	32
55. Zn (0001)	105	−196	Crystal cleavage (N <sub>2</sub> )	32
56. Fe (100)	1360	−250	Crystal cleavage (H <sub>2</sub> )	32
57. C (graphite) (001)	73	25	Free energy of immersion	45
58. C (graphite) (001)	1750	25	Crystal cleavage in UHV	46
59. C (graphite) (001)	2500	25	Theoretical calculation	46
60. Mica (Muscovite)	10 250	25	Crystal cleavage in UHV	47
61. Mica (Muscovite)	300	25	Crystal cleavage in air	48
62. Silver (100)	900	1140	Creep of thin wires	49
$W = \pi r \gamma - \frac{n}{l} \pi r^2 \gamma$				
63. Gold (100)	1400	1300	Creep of thin wires	50
$W = \pi r \gamma - \frac{n}{l} \pi r^2 \gamma$				
64. Gold (111)	1510	-----	-----	28
65. Copper (100)	1430	1050	Creep of thin wires	29
66. Copper (111)	1670	1050	-----	28
67. Silver (100)	1150	900	Calculated from eq. (9)	28
68. Silver (111)	1186	900	Calculated from eq. (9)	28
69. Gold (100)	1396	1300	Calculated from eq. (9)	28
70. Gold (111)	1438	1300	Calculated from eq. (9)	28
71. Copper (100)	1509	1050	Calculated from eq. (9)	28
72. Copper (111)	1560	1050	Calculated from eq. (9)	28
73. $\gamma$ Iron	2150	1360	Creep of thin wires	51
74. $\delta$ Iron	1950	1400	Creep of thin wires	51
75. Nickel	1850	1250	Creep of thin wires	52
76. Nickel (100)	1821	1219	Surface relaxation	31
77. Nickel (110)	1900	1219	Surface relaxation	31
78. Niobium	2100	2250	Creep of fine wires	53
79. Platinum	2340	1310	Surface relaxation	54
80. Tungsten	2900	1727	Field emission	30
Critical Surface Tension for Wetting of Low-Energy Solids				
81. PTFE	18.5	20	Contact angle measurements	55
82. Polyethylene	31	20	Contact angle measurements	55
83. Polystyrene	33	20	Contact angle measurements	55
84. Poly vinyl alcohol	37	20	Contact angle measurements	55
85. Polymethyl methacrylate	39	20	Contact angle measurements	55

The dispersion forces appear in the corrections that must be made to the ideal gas law. For an ideal gas the quotient  $PV/T$  is a constant. But real gases are more compressible at low density and less so at high pressures. The first effect is the result of mutual attractions between gas molecules that arise from the dispersion forces. The second effect occurs because the gas molecules occupy a finite incompressible portion of the volume, and the molecules will strongly repel each other if they come into contact.

London calculated the dispersion forces from reasoning based on the quantum mechanics (refs. 59 and 60). In this model, an atom or molecule having a time-average spherically symmetrical charge distribution will at any instant have a momentary dipole moment due to continuous fluctuations in the charge distribution. This fluctuating dipole can induce dipoles in neighboring molecules, and an interaction energy results:

$$E = C_L/r^6 \quad (10)$$

$C_L$  is a constant from London's calculation equal to  $(3/4)\alpha^2 I$  with  $\alpha$  the polarizability of the molecule and  $I$  its ionization energy.

This calculation does not apply for fairly large distances of separation. The dipole fluctuates according to certain characteristic frequencies. The neighboring molecule's capability to respond to this polarizing electromagnetic field must be limited by the speed of light. At large separations the neighboring molecule cannot respond before the charge fluctuation in the polarizing molecule has moved on to some new configuration. As a result the dipoles are out of phase, and the energy of interaction is reduced. The correct functional relationship now is (refs. 60 and 61)

$$e = -C_r/r^7 \quad (11)$$

where a new constant,  $C_r$ , is now used for this retarded dispersion force.

Formulas for the force of attraction between two macroscopic bodies have been calculated based on the normal and retarded dispersion force pictures. Table 2 gives the results for parallel plates and a sphere on a flat in dynes/cm<sup>2</sup>.

Attempts to measure these forces by sensitive techniques have been encumbered by formidable experimental difficulties. In any event, approach closer than a few thousand angstroms was not possible, and for these separations the results seemed to be in harmony with the retarded dispersion force calculation. Bowden (refs. 62 and 63) has attempted to measure the forces between sheets of mica in a cleavage. The shape of the two leaves leading into the crack should be determined by the mechanical properties. But near the point of bifurcation where the surfaces are close together, the shape should be altered by the surface forces. This shape was measured using multiple interference techniques. The results indicated that in the range 20 to 50 Å, the force fell off as the inverse

TABLE 2.—*Dispersion Forces on Macroscopic Bodies (dynes/cm<sup>2</sup>)\**

Geometry	Normal	Retarded
Parallel plates	$F = \frac{A}{6\pi h^3}$	$F = \frac{B}{h^4}$
Hemisphere on flat	$F = \frac{AR}{6h^2}$	$F = \frac{2\pi BR}{3h^3}$

\*  $A$  and  $B$  are constants;  $h$  is separation;  $R$  is radius of sphere.

cube of the separation as predicted by London's theory for unretarded forces between parallel plates.

Tabor and Winterton (ref. 64), again using mica cleavage surfaces, were able to measure the attractive force over the range of separations from 50 to 300Å. The functional dependence of the attractive force on separation changed in the region around 150Å. Below this separation the data fit the theory for unretarded dispersion forces; above this separation the retarded forces held sway.

The magnitudes of the forces between macroscopic surfaces at several separations are given in table 3. The forces become appreciable at separations somewhere below 1000Å.

It appears that it is not necessary to assign any peculiar forces to surfaces to account for their properties. Interactions between macroscopic bodies as well as colloidal particles can be understood in terms of familiar force fields. The role played by the dispersion forces can be particularly important where surfaces and disperse systems are being considered.

The effects of surface chemical constitution on the surface energy of some low energy surfaces have been clarified in some refined experiments of Zisman (ref. 55) and his coworkers. A straight line results when  $\cos \theta$  ( $\theta$  is the contact angle) is plotted as a function of liquid surface tension for a series of liquids, all of the same homologous series on a given low energy surface. The intercept at  $\cos \theta = 1$  (at which point the surface is

TABLE 3.—*Measured Dispersion Forces Between Macroscopic Surfaces*

$h$ , Å	Unit Forces, dynes/cm <sup>2</sup>	Ref.
2000	1.0	30
1000	10.0	30
100	$1.5 \times 10^6$	63
20	$3 \times 10^6$	63



completely wet) is the critical surface tension for wetting, and this value is reasonably independent of the series of compounds which is used for the measurements. Any liquid with a critical surface tension for wetting less than this value would be expected to spread over the surface. Moreover, the values of  $\gamma_c$  are shown to be determined by the chemical composition of the surface in a quite regular fashion. In table 4, the effect of replacing fluorine in the surface by hydrogen is shown.

An alternative empirical view of solid surface interaction liquids has been given by Fowkes (ref. 65). It is postulated that force interactions across an interface can be broken down into component parts (i.e., hydrogen bonding, dispersion, metallic, etc.). The interfacial energy for two contacting materials will arise only from those sources of surface forces common to the two materials. For dispersion force interaction, a plot of  $\cos \theta$  as a function of  $\sqrt{\gamma_L^d/\gamma_L}$  should be linear with a slope of  $2\sqrt{\gamma_s^d}$ . This provides a method for estimating the dispersion force component to the surface energy of a solid. The values determined for a number of low energy surfaces yield values close but not identical with the values of critical surface tension for wetting.

#### ADSORPTION AND SURFACE THERMODYNAMICS

The two general types of adsorption have already been discussed. It is very difficult to differentiate between chemical and physical adsorption in a very clear-cut manner. It is perhaps best to consider them in terms of the adsorption mechanisms involved; that is, the sources of interaction energy, rather than in terms of the relative amounts of energy involved or the assumed reversibility or irreversibility of the processes. Chemical adsorption involves the formation of a chemical compound at the surface with a composition different from the bulk phase. These surface compounds are no different than those otherwise encountered in chemistry. Chemical adsorption is an important phase in the mechanism of contact

TABLE 4.—*Dependence of Critical Surface Tension for Wetting of Surface Constitution*

Surface group	$c$ , dynes/cm
—CF <sub>3</sub>	6
—CF <sub>2</sub> H	15
—CF <sub>3</sub> and —CF <sub>2</sub> —	17
—CF <sub>2</sub> —	18
—CH <sub>2</sub> —CF <sub>3</sub>	20
—CF <sub>2</sub> —CFH—	22
—CF <sub>2</sub> —CH <sub>2</sub> —	25
—CFH—CH <sub>2</sub> —	28

catalysis (ref. 66). Nascent metal surfaces formed by wear processes are extremely reactive chemically, and a monolayer of chemisorbed species should form on such surfaces practically instantaneously under normal circumstances (ref. 67). Most metal surfaces that we deal with already have a chemisorbed surface layer leading to a much reduced surface energy. Further chemical reaction will be slow at ordinary temperatures. However, physical adsorption can proceed. Physically adsorbed gases on surfaces can be fairly mobile and multilayers of physisorbed gases can be present depending on conditions. Such films can play important roles in the lubrication of concentrated contacts.

Physical adsorption has been a subject of considerable theoretical concern. Over a limited pressure range, the amount of adsorption can often be fit by the empirical Freundlich adsorption isotherm

$$X = kp^{1/n} \quad (12)$$

Here  $X$  is the mass of gas adsorbed by 1 mg of the adsorbing solid,  $p$  is the pressure, and  $k$  and  $n$  are constants. Langmuir treated the problem theoretically by considering a unimolecular layer in which the processes of adsorption due to gas molecules striking the surface and evaporation from the film are in dynamic equilibrium (ref. 68). The result is the Langmuir adsorption isotherm

$$\theta = \frac{Ap}{1 + Bp} \quad (13)$$

where  $\theta$  is the fraction of the surface covered by adsorbed molecules,  $p$  is the partial pressure of the adsorbing species in the gas phase, and  $A$  and  $B$  are constants. For convenience in handling experimental data, the Langmuir isotherm is generally put into some sort of linear form. A useful arrangement is

$$\frac{p}{v} = \frac{1}{bv_m} + \frac{p}{v_m} \quad (14)$$

Here  $v$  is the volume of gas adsorbed per gram of adsorbant,  $v_m$  is the volume of gas adsorbed for monolayer coverage, and  $b$  is a constant given by

$$b = N\sigma_o\tau_o' \left( \frac{1}{2\pi RTM} \right)^{1/2} \exp \frac{Q}{RT} \quad (15)$$

where  $N$  is Avogadro's number,  $\sigma_o$  is the actual area of an adsorbate molecule,  $\tau_o'$  is an interaction time between the adsorbate molecule and the surface (of the order of  $10^{-13}$  sec),  $R$  is the gas constant,  $T$  is the absolute temperature,  $M$  is the molecular weight of the gas, and  $Q$  is the heat of adsorption. It is seen that a plot of  $p/v$  as a function of  $p$

should be linear with a slope of  $1/v_m$  and an intercept of  $1/bv_m$ . So the adsorption isotherm can be used to reveal the heat of adsorption and the area of the adsorbant if certain assumptions are made concerning the dimensions of the adsorbate molecule. The determination of surface areas has made extensive use of the interpretation of adsorption isotherms. The concept of the surface area of an irregular solid surface possesses considerable ambiguity. The surface area measured is really defined in terms of the method used to measure it. Obviously in adsorption studies, the surface area measured will show some dependence on the gas used and will be dependent on such factors as molecular size and shape.

Langmuir originally derived his adsorption isotherm using a kinetic argument in which the rate of the gas striking the surface and being retained became equilibrated with the rate of gas evaporating from the surface. It has been pointed out that a number of serious assumptions are implied in this derivation (ref. 69). It is seen that one of the constants in the equation involves the energy of adsorption  $Q$ ; and for  $b$  to be a true constant,  $Q$  must not vary with the amount of surface coverage. This amounts to assuming that the solid surface is homogeneous with all adsorption sites possessing the same interaction energy with the adsorbate molecules and with no interaction between adsorbate molecules on neighboring sites. The experimental evidence reveals that  $Q$  usually does change with the amount of surface coverage, and in many instances this effect cannot be ignored.

A second serious assumption is that the adsorption is localized on the sites. A gas molecule is presumed to be adsorbed when it strikes an unoccupied site and remains there without the possibility of migrating over the surface to other unoccupied sites. A statistical analysis reveals that this amounts to making a restrictive assumption concerning the entropy of adsorption. Again experimental evidence suggests that adsorbate molecules on the surface are mobile. Indeed, there is evidence that a gas molecule can physically adsorb on a covered surface and then migrate over this surface to find an unoccupied site.

This last observation points to yet another deficiency in the theoretical treatment. The use of  $v_m$  contains the assumption that adsorption cannot continue beyond a single monolayer coverage. This difficulty has been removed by a refinement of the Langmuir treatment (the BET equation (ref. 70) which allows for multilayer formation), and it is in this form that the theory is used in the interpretation of isotherms to derive surface area data.

Many other approaches have been developed to derive equations that account for adsorption data. These have been based on potential theory as well as on certain forms of the equation of state for the adsorbed gas. In many instances, a good fit to experimental data is obtained as well as reasonable values for surface areas. It appears that as the detailed atomic



processes involved in adsorption become better understood, more rigorous theoretical treatments will be possible. However, their probable complexity makes it unlikely that the Langmuir description will be quickly supplanted.

Adsorbed surface layers are assigned an important role in boundary lubrication. The fact that a substance adsorbs readily on a surface at the solid liquid interface is no assurance that an effective lubricating layer has been formed. Work by Zisman and his coworkers has indicated that, for long chain molecules adsorbed on metal and glass surfaces, friction and wear on the adsorbed film are low only if it assumes the structure of a two-dimensional solid as indicated by contact angle measurements with methylene iodide (ref. 71). This implies virtually perfect alignment of the molecules at the interface. How much of this picture, based on a somewhat idealized experiment, applies to the practical situation of metal surface subject to wear in the presence of a lubricant containing a boundary additive is the subject of some speculation. The formation of the two-dimensional crystalline interface would appear to be difficult from a number of different viewpoints. Surface irregularities and imperfections would introduce geometric difficulties for packing of the long-chain additives which attempt to align normal to the surface. In the adsorption process, the additive must compete with other agents present in the total lubricant system, and a pure adsorbed layer would appear to be a remote possibility (ref. 72). Sliding is a kinetic process and any mechanism that is to be effective in reducing friction and wear must be rapid relative to the times involved in sliding. Adsorption is not instantaneous. There unfortunately is not an abundance of data on the rates of adsorption from solution on smooth solid surfaces. It is reported that, in adsorption of stearic acid from benzene on iron powder, it took 4 hours for equilibrium to be reached though 90 percent of the stearic acid reached the surface within 5 minutes (ref. 73). Data for rates of adsorption at the liquid-vapor surface are more abundant. If fresh surface of a pure liquid is generated, the surface tension rapidly achieves a steady value in less than 0.01 second. Any time required for equilibration probably involves re-orientation of molecules already automatically present at the interface. For solutions the situation is more complex, and seconds to minutes are required for the surface tension to equilibrate in a freshly formed surface from a solution of a long chain alcohol or acid (ref. 74). The first step of this process may involve diffusion of molecules from the interior to the subsurface region followed by final entry into the interface. The first molecules to arrive enter the interface readily and probably assume random orientations with very inefficient packing. Thus late arrivals can only achieve entry into the interface if there is large scale re-orientation of molecular chains. This can be a slow process. Free energies of activation for adsorption of long chain dibasic acids at the

liquid-vapor interface can be 14 to 18 kcal/mole (ref. 75). This is fairly high for something which is generally regarded to be a physical process.

Extrapolation of these results to the solid-liquid interface where lubrication phenomena occur is of course dangerous. However, it seems reasonable to call into question any idealized picture of a boundary lubricant film as a perfectly oriented, pure monolayer having the properties of a perfectly crystalline two-dimensional solid. Certainly the time required to form such a film must be long relative to the time made available by the sliding process under conditions encountered in normal engineering practice. There is still much to be learned concerning the mechanism of action of a boundary lubricant and the details of composition of boundary lubricating layers, their structure, and the mechanisms of formation under actual sliding conditions.

A relationship between the ability of a boundary lubricant to prevent wear and its heat of adsorption has been proposed by Rowe (ref. 76). The author develops a wear equation in which the volume of wear per unit distance of sliding is related to the wear coefficient for a metal to metal contact and the fraction of the surface,  $\alpha$ , covered by the boundary lubricant molecules. The value of  $\alpha$  is related to the residence time for an adsorbed molecule on a surface site. The final equation is

$$\frac{V}{d} = \frac{k_m(1+3f^2)^{1/2}XW}{t_o P_m U} e^{-E/RT_s} \quad (16)$$

where  $V$  is the volume of wear,  $d$  is the sliding distance,  $k_m$  is the wear coefficient for metal to metal contacts,  $f$  is the coefficient of friction,  $X$  is the diameter of the adsorbed molecule,  $W$  is the load,  $t_o$  is the fundamental vibration frequency for the molecule in the adsorbed state,  $P_m$  is the yield pressure of the metal,  $U$  is the sliding velocity,  $R$  is the gas constant,  $T_s$  is the surface temperature, and  $E$  is the energy of adsorption and is referred to as the heat of adsorption in the author's discussion. Because of the relatively low interaction energies for the physically adsorbed species, the surface films will remain unpenetrated only under relatively mild conditions of loading and surface speed. Therefore, the analysis is restricted to a description of mild wear conditions. When data on the wear of a copper pin against a steel disc in cetane were analyzed according to the theory, good correlation was observed with the calculated values for heat of adsorption values estimated from transition type friction and wear results. The analysis has also been applied to the vapor lubrication of graphite (ref. 77) with good correlation being found between the heat of adsorption of the vapor on graphite and the minimum pressure required to prevent wear.

Another attempt to relate friction and wear phenomena to surface energetics involves application of the work of adhesion concept (ref. 78). The concept has been applied to predict the size of loose wear particles



formed between sliding metal surfaces. Here the model system pictures the potential wear particle as a hemisphere in contact with a flat. The particle will separate if the stored elastic energy exceeds the work of adhesion at the contact. Then a diameter for a typical wear particle should be

$$d = \frac{60\,000 W_{AB}}{p} \quad (17)$$

Considering the assumptions in the derivation and the difficulties in designating appropriate surface energies for practical sliding couples, agreement with experiment is reasonable. A difficulty in the analysis is that it predicts a single particle diameter, and in practice a range of particle sizes is observed. In the test of the expression an average particle was defined such that half of the weight in the wear fragments was in particles of larger size. Despite the arbitrariness of this procedure, the equation appears to predict the particle size within a factor of three for like metal pairs.

It appears that equilibrium thermodynamic approaches to predicting friction and wear events may have serious limitations. Many asperity contacts may occur without producing a wear particle or material transfer. We cannot as yet identify those primary events that trigger the wear process. It seems that a rational explanation must relate back to a more complete description of the physical and chemical conditions of the contacting surfaces along lines previously discussed in this review. Such a description is not yet possible, but we seem to be approaching it at an accelerating rate.

#### DISCUSSION

**D. H. Buckley** (Lewis Research Center, Cleveland, Ohio)

It was a real pleasure to read Dr. Haltner's paper because he has stressed the importance of our gaining a better understanding of surfaces from an atomic point of view. As Dr. Haltner has indicated, the nature of the surface and the interaction of surfaces with the environment are extremely important in the understanding of adhesion, friction, and wear. The availability of new research tools to assist in our analysis of surfaces have put the ability to better understand surfaces and interfaces more closely within our reach.

In our laboratory we are presently using two of the tools Dr. Haltner referred to as a part of our adhesion studies, namely, the field ion emission microscope and LEED (low-energy electron diffraction). I would like to mention one tool that can be extremely useful for those of us working in the field of lubrication which Dr. Haltner did not discuss. This is the Auger emission spectrometer. Auger emission spectrometry enables one to determine which elements are present on a surface to coverages of less than 0.1 of a monolayer, and this sensitivity should be improved shortly.



Auger involves the measurement of the energy associated with the emission of secondary electrons from the electron shells of atoms. It is especially good in the analysis of the lower-atomic-number elements, and both single and polycrystalline surfaces can be studied. Elements such as carbon, sulfur, oxygen, fluorine, and chlorine are easily detected on a surface.

We have incorporated Auger emission spectrometry into a LEED system. The coupling of these two devices permits both a structural and elemental analysis of a surface in the same system. Thus one can obtain not only a structural analysis of surface films present in fractions of a monolayer, but also identify them elementally. With such tools it should be possible to answer some of the questions that have eluded us in the past, for example, the exact role and interaction of sulfur in EP additives with ferrous surfaces and the influence of oxygen on that interaction.

The influence of the presence of small amounts of foreign agents in liquids and solids referred to by Dr. Haltner cannot be too strongly stressed. In sliding friction studies with calcium fluoride, we have found that concentrations of dimethylsulfoxide in water of  $10^{-8}$  to  $10^{-10}$  molar were sufficient to influence friction and deformation (Rehbinder effect) in sliding. This certainly is in the range of the impurity level of many of our commonly used lubricants.

It is important to recognize that the impurities may come not only from the lubricant, but also from the material to be lubricated. In LEED and Auger studies with copper-aluminum alloys, we have found that concentrations of aluminum in bulk copper of 1.0 atomic percent are sufficient to alter adhesion of surfaces markedly (fivefold increase with 1 percent aluminum present in the copper).<sup>\*</sup> Auger analysis indicated a tenfold greater concentration of aluminum on the surface than in the bulk due to equilibrium segregation of aluminum on the surface. This equilibrium segregation occurs at temperatures as low as 200° C, well within the temperatures that may occur on surfaces in many lubricated devices. Such segregation has been seen for sulfur on stainless steel, niobium, and uranium where bulk sulfur concentrations were at the impurity level (ref. 79). In the case of uranium, only 20 ppm was present in the bulk.

In table 1 Dr. Haltner presents surface energy values for a number of materials. The reader should be cautioned that these values may suffer from some of the infirmities referred to by Dr. Haltner in his paper, namely, the questions of purity and cleanliness. For example, for the copper (111) surface, energy values in the literature can be found that

---

<sup>\*</sup> Buckley, D. H.: Effect of Various Properties of FCC Metals on Their Adhesion as Studied with LEED. To be published in J. of Adhesion.

range from the 1600 ergs/cm<sup>2</sup> presented in the author's table 1 to 2900 ergs/cm<sup>2</sup>. In this regard, some of the older values may be particularly suspect since high-purity materials were not available and surface analysis techniques in use today were not employed to define surface conditions.

The author discusses adsorption and, as most of us tend to do, distinguishes between physical and chemisorption. While physical adsorption has been the subject of theoretical concern and is fairly well understood, its exact role in lubrication is not yet clear. Some very fine studies on the influence of heats of adsorption in lubrication have been conducted by Rowe (refs. 76 and 77) and Groszek (ref. 80). No distinction between the two types of adsorption are made in these works. The Rowe analysis appears to apply to physically adsorbed species, yet the model does not appear to distinguish between the two types of adsorption. The heats of adsorption range from 5 to 50 kcal/mole in the studies by these investigators. Chemisorption can occur with heats of adsorption as low as 3 kcal/mole (ref. 66). While heats of adsorption alone are not the only criterion to be considered, they do give a good indication of the types of adsorption.

Examination of texts on adsorption indicate five factors that may be used to distinguish between the two types of adsorption (refs. 66 and 69). The first factor is the magnitude of the heat of adsorption because chemical bonds are normally stronger than physical forces of attraction. The second factor is the temperature range over which adsorption proceeds. Since physical adsorption is related to liquefaction, it usually occurs at temperatures near or below the boiling point of the adsorbate. Chemisorption, as is well known, can occur at temperatures well above the boiling point. The third point is that chemisorption may require an activation energy while physical adsorption requires none. Fourth, chemisorption has some specificity while physical adsorption does not. Fifth and last is that chemisorption is a monolayer process; that is, chemisorption ceases when the adsorbate can no longer make direct contact with a surface. Physically adsorbed layers, on the other hand, may be many molecular layers thick. Anything on a surface beyond a monolayer must therefore be physically adsorbed.

The distinction between these two types of adsorption is not merely of academic interest but has real significance to the behavior of solid surfaces in contact. Chemisorbed films may inhibit adhesion of two surfaces in contact at temperatures and pressures where physically adsorbed film do not exist. Chemisorbed films may inhibit or alter the reactivity of a surface with other surface-active species. Further, the effectiveness of chemisorbed films in influencing adhesion may be related in part to the fact that electron transfer takes place in chemisorption which alters the very character of the surficial layer. This does not occur with physically adsorbed layers which, as Dr. Haltner has pointed out, are relatively



mobile. It would be of interest to know just what effect a truly physically adsorbed film, say argon on a clean metal surface, has on measured adhesive forces between two clean metal surfaces.

In regard to the work of adhesion-wear particle diameter concept, one wonders if any correlations here are not more accidental than real. Surface energy is a surface property that we know has some real relation to adhesion. The higher surface energy planes of a metal usually exhibit stronger adhesion. The generation of a wear particle is, however, more than simply the result of interfacial adhesions. Whenever a wear particle is generated, subsurface fracture, either ductile or brittle, must have occurred. If fracture always occurred at the interface where adhesion of the two surfaces takes place, a wear particle would never be generated. One could still, however, talk of energy of adhesion.

While adhesion plays a part in the generation of a wear particle, there are many subsurface factors that may play a role as important or more important in determining the nature and size of a wear particle. Concentration of vacancies, dislocations, and interstitials will exert an influence. Further, those factors that are known to give rise to the formation of fracture, such as dislocation coalescence at slip plane interaction and dislocation pile-ups at pinning elements, certainly have some influence on the fracture path through a metal.

In some simple adhesion experiments with approximately 12 face-centered cubic metal pairs, we found that in the adhesion process the cohesively weaker of the two metals in contact always transferred to the stronger. The adhesive junction was stronger than the cohesive bonds of the weaker of the two materials, and fracture occurred subsurface in the weaker of the two metals. The adhesive force and amount of transfer could be related to lattice mismatch. The closer the lattice matching, the stronger the adhesive bond and the greater the material transfer. With an increase in lattice mismatch there is an increase in lattice coherency strain and an increase in the concentration of subsurface misfit dislocations in the cohesively weaker material. It was these subsurface imperfections that gave rise to the zone of greatest weakness in the surficial zone and the region of subsurface fracture.

#### LECTURER'S CLOSURE

*Editor's Note*—Mr. Buckley's informative and provocative contribution was duly forwarded to Dr. Haltner for comment. Dr. Haltner has not, however, submitted a closure for publication.

#### REFERENCES

1. HARDY, W. B.: *Collected Works*. Cambridge Univ. Press, 1936.
2. LING, F. F.; KLAUS, E. E.; AND FEIN, R. S., eds.: *Boundary Lubrication, An Appraisal of World Literature*. ASME, 1969.



3. GODFREY, D.: Boundary Lubrication. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, pp. 335-384.
4. CAMPBELL, W. E.: Discussion on Boundary Lubrication by D. Godfrey. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, pp. 353-358.
5. DOWSON, D.: Transition to Boundary Lubrication from Elastohydrodynamic Lubrication. Boundary Lubrication, An Appraisal of World Literature, ch. 11, F. F. Ling, E. E. Klaus, and R. S. Fein, eds., ASME, 1969.
6. DOWSON, D.; AND HIGGINSON, G. R.: Elastohydrodynamic Lubrication. Pergamon Press, 1966.
7. CAMERON, A.: Principles of Lubrication. Chs. 8 and 10, John Wiley & Sons, Inc., 1966.
8. FEIN, R. S.; AND KREUZ, K. L.: Discussion on Boundary Lubrication by D. Godfrey. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, pp. 358-376.
9. WILLIAMSON, J. B. P.: Topography of Solid Surfaces. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, pp. 85-142.
10. KRAMER, I. R.; AND DEMER, L. J.: Effects of Environment on Mechanical Properties of Metals. Prog. in Materials Sci., vol. 9, 1961, pp. 131-199.
11. KRAMER, I. R.: Effect of Surfaces on Mechanical Behavior of Metals. Fundamental Phenomena in the Materials Sciences, vol. 3, Surface Phenomena, L. J. Bonis, P. L. deBruyn, and J. J. Duga, eds., Plenum Press, 1966.
12. DOWSON, D.; AND HIGGINSON, G. R.: New Roller Bearing Lubrication Formula. Engineering, vol. 192, 1961, pp. 158-159.
13. BOWDEN, F. P.; AND TABOR, D.: The Friction and Lubrication of Solids. pt. I, chs. IX-X, 1953; pt. II, chs. XVIII-XIX, 1964, Oxford Univ. Press (London).
14. CAMPBELL, W. E.: Boundary Lubrication. Boundary Lubrication, An Appraisal of World Literature, ch. 6, F. F. Ling, E. E. Klaus, and R. S. Fein, eds., ASME, 1969.
15. HALTNER, A. J.: The Application of Ultrahigh Vacuum Techniques to Studies of Friction and Wear. Trans. Tenth Nat. Vacuum Symp., Amer. Vac. Soc., 1963, pp. 14-20.
16. ROBERTS, R. W.: Clean Surfaces, Their Preparation and Characterization. Cold Welding of Materials in Space Environments, STP 431, ASTM, 1967, pp. 20-66.
17. RICHTMYER, F. K.; AND KENNARD, E. H.: Introduction to Modern Physics. Fourth ed., ch. III, McGraw-Hill Book Co., Inc., 1947.
18. MACRAE, A. U.: Techniques for Studying Clean Surfaces. Surfaces and Interfaces. I: Chemical and Physical Characteristics, J. J. Burke, N. L. Reed, and V. Weiss, eds., Syracuse Univ. Press, 1967, pp. 29-52.
19. GATOS, H. C.: Structure of Surfaces and Their Interactions. Interdisciplinary Approach to Friction and Wear, P. M. Ku, ed., NASA SP-181, 1968, pp. 7-84.
20. GOMER, R.: Field Emission and Field Ionization. Harvard Univ. Press, 1961.
21. MULLER, E. W.: Field Ion Microscopy of Surface Structures on an Atomic Scale. Properties of Surfaces, STP 340, ASTM, 1963, pp. 80-97.
22. MULLER, E. W.; NAKAMURA, S.; NISHIKAWA, O.; AND McLANE, S. B.: Gas-Surface Interactions and Field-Ion Microscopy of Non-refractory Metals. J. Appl. Phys., vol. 36, 1965, p. 2496.
23. MULLER, E. W.; AND NISHIKAWA, O.: Atomic Surface Structure of the Common Transition Metals and the Effect of Adhesion as Seen by Field Ion Microscopy. Adhesion or Cold Welding of Materials in Space Environments, STP 431, ASTM, 1967, pp. 67-87.
24. OGILVIE, R. E.: Quantitative Electron Microprobe Analysis. X-Ray and Electron

- Methods of Analysis, ch. III, H. van Olphen and W. Parrish, eds., Plenum Press, 1968.
25. GUGGENHEIM, E. A.: The Principle of Corresponding States. *J. Chem. Phys.*, vol. 13, 1945, p. 253.
  26. JURA, G.; AND GARLAND, C. W.: The Experimental Determination of Surface Tension of Magnesium Oxide. *J. Am. Chem. Soc.*, vol. 74, 1952, p. 6033.
  27. PARTINGTON, J. R.: An Advanced Treatise on Physical Chemistry, vol. 3, The Properties of Solids, Longmans, Green and Co. (London), 1952, pp. 242-247.
  28. BONDI, A.: The Spreading of Liquid Metals on Solid Surfaces. *Chem. Rev.*, vol. 52, 1953, p. 417.
  29. UDIN, H.: Measurement of Solid: Gas and Solid: Liquid Interfacial Energies. *Metal Interfaces*, ASM, 1952, pp. 114-150.
  30. BARBOUR, J. P.; CHARBONNIER, F. M.; DOLAN, W. W.; DYKE, W. P.; MARTIN, E. E.; AND TROLAN, J. K.: Determination of the Surface Tension and Surface Migration Constants for Tungsten. *Phys. Rev.*, vol. 117, 1960, p. 1452-1459.
  31. BLAKELY, J. M.; AND MAIYA, P. S.: Surface Energies from Transport Measurements. *Surface and Interfaces. I: Chemical and Physical Characteristics*, J. J. Burke, N. L. Reed, and V. Weiss, eds., Syracuse Univ. Press, 1967, pp. 325-2331.
  32. GILMAN, J. J.: Surface Energies of Crystals. *J. Appl. Phys.*, vol. 31, 1960, p. 2208.
  33. ANON.: Handbook of Chemistry and Physics. 45th ed., The Chemical Rubber Co., Cleveland, Ohio, 1964.
  34. GAINES, G. L.: The Surface Tension of Polymer Solutions. I: Solutions of Poly(dimethyl siloxanes). General Electric Co. Res. and Dev. Ctr. Rpt. 69-C-059, 1969.
  35. LEGRAND, D. G.; AND GAINES, G. L.: The Molecular Weight Dependence of Polymer Surface Tension. General Electric Co. Res. and Dev. Ctr. Rpt. 69-C-055, 1969.
  36. SHARTSIS, L.; SPINNER, S.; AND SMOCK, A. W.: Surface Tension of Compositions in the Systems  $\text{PbO-B}_2\text{O}_3$  and  $\text{PbO-SiO}_2$ . *J. Am. Ceram. Soc.*, vol. 31, 1948, p. 23.
  37. LANDOLT-BORNSTEIN: *Physikalish Chemische Tabellen*. Fifth ed., vol. 1, Springer Verlag (Berlin), 1923.
  38. LYON, R. N., ed.: *Liquid Metals Handbook*. U.S. Navy Office of Naval Res., Second ed., 1952.
  39. HARKINS, W. D.: *Physical Chemistry of Surfaces*. Reinhold Publ. Corp., 1952.
  40. LANDOLT-BORNSTEIN: *Physikalish Chemische Tabellen*. Fifth ed., vol. IIIA, Springer Verlag (Berlin), 1923.
  41. ADAM, N. K.: *Physics and Chemistry of Surfaces*. Oxford Univ. Press (London), 1941.
  42. GOOD, R. J.: Theory for the Estimation of Surface and Interfacial Energies. Contact Angle, Wettability and Adhesion, *Advances in Chemistry Series 43*, ACS, 1964, p. 74.
  43. MELROSE, J. C.: Evidence for Solid-Fluid Interfacial Tensions from Contact Angles. Contact Angle, Wettability and Adhesion, *Advances in Chemistry Series 43*, ACS, 1964, p. 158.
  44. MARK, H. F.: Future Trends for Improvement of Cohesive and Adhesive Strength of Polymers. *Adhesion and Cohesion*, P. Weiss, ed., Elsevier Publ., 1962, pp. 240-269.
  45. GOOD, R. J.; GIRIFALCO, L. A.; AND KRAUS, G.: A Theory for Estimation of Interfacial Energies. II. Application to Surface Thermodynamics of Teflon and Graphite. *J. Phys. Chem.*, vol. 62, 1958, p. 1418.
  46. BRYANT, P. J.; GUTSHALL, P. L.; AND TAYLOR, L. H.: A Study of Mechanisms of



- Graphite Friction and Wear. Mechanisms of Solid Friction, P. J. Bryant, M. Lavik, and G. Salomon, eds., Elsevier Publ., 1964, p. 1118.
47. BRYANT, P. J.: Cohesion of Clean Surfaces and the Effect of Adsorbed Gases. Trans. Ninth Nat. Vac. Symp., Amer. Vac. Soc., 1962, p. 311.
48. BRYANT, P. J.; TAYLOR, L. H.; AND GUTSHALL, P. L.: Cleavage Studies of Lamellar Solids in Various Gas Environments. Trans. Tenth Nat. Vac. Symp., Amer. Vac. Soc., 1963, p. 21.
49. BUTTNER, F. H.; FUNK, E. R.; AND UDIN, H. J.: Adsorption of Oxygen on Silver. J. Phys. Chem., vol. 56, 1952, p. 657.
50. BUTTNER, F. H.; UDIN, H.; AND WULFF, J.: Surface Tension of Solid Gold. AIME J. of Metals, vol. 3, 1951, pp. 1206-1209.
51. PRICE, A. T.; HOLL, H. A.; AND GREENOUGH, A. P.: The Surface Energy and Self-Diffusion Coefficient of Solid Iron Above 1350° C. Acta. Met., vol. 12, 1964, p. 49.
52. HAYWARD, E. R.; AND GREENOUGH, A. P.: The Surface Energy of Solid Nickel. J. Inst. Metals, vol. 88, 1960, p. 217.
53. RADCLIFFE, S. V.: The Surface Energy of Solid Niobium. J. Less-Common Metals, vol. 3, 1960, p. 360.
54. BLAKELY, J. M.; AND MYKURA, H.: Surface Self-Diffusion and Surface Energy Measurements in Platinum. Acta. Met., vol. 10, 1962, p. 565.
55. ZISMAN, W. A.: Relation of the Equilibrium Contact Angle to Liquid and Solid Constitution. Contact Angle, Wettability and Adhesion, Advances in Chemistry Series 43, ACS, 1964, pp. 1-51.
56. LENNARD-JONES, J. E.: Processes of Adsorption and Diffusion on Solid Surfaces. Trans. Faraday Soc., vol. 28, 1932, pp. 333-359.
57. SAVAGE, R. H.: Graphite Lubrication. J. Appl. Phys., vol. 32, 1961, p. 4.
58. EHRLICH, G.: Kinetic and Experimental Basis of Flash Desorption. J. Appl. Phys., vol. 32, 1961, p. 4.
59. PAULING, L.; AND WILSON, E. B.: Introduction to Quantum Mechanics. Ch. XIV, McGraw-Hill Book Co., Inc., 1935.
60. DEBYE, P. J. W.: Interatomic and Intermolecular Forces in Adhesion and Cohesion. Adhesion and Cohesion, P. Weiss, ed., Elsevier Publ., 1962, pp. 1-17.
61. LIFSHITS, E. M.: The Theory of Molecular Attraction Forces Between Solid Bodies. Zhur. Eksptl. i Teoret. Fiz., vol. 29, 1955, pp. 94-110. (Chem. Abs., vol. 49, 1955.)
62. BOWDEN, F. P.: The Adhesion of Metals and the Influence of Surface Contamination and Topography. Adhesion and Cohesion, P. Weiss, ed., Elsevier Publ., 1962, pp. 121-145.
63. BOWDEN, F. P.: The Nature and Topography of Solid Surfaces and the Study of Van der Waals' Forces in Their Immediate Vicinity. The Surface Decomposition of Solids. Fundamentals of Gas-Surface Interactions, H. Saltsburg, J. N. Smith, Jr., and M. Rogers, eds., Academic Press, 1967, pp. 1-24.
64. TABOR, D.; AND WINTERTON, R. H. S.: Surface Forces: Direct Measurement of Normal and Retarded Van der Waals' Forces. Nature, vol. 219, 1968, pp. 1120-1121.
65. FOWKES, F. M.: Intermolecular and Interatomic Forces at Interfaces. Surfaces and Interfaces. I: Chemical and Physical Characteristics, J. J. Burke, N. L. Reed, and V. Weiss, eds., Syracuse Univ. Press, 1967, pp. 197-224.
66. HAYWARD, D. O.; AND TRAPNELL, B. M. W.: Chemisorption. Second ed., Butterworths, London, 1964.
67. ROBERTS, R. W.: Adsorption and Decomposition of Hydrocarbons on Clean Metal Films. Brit. J. Appl. Phys., vol. 14, 1963, p. 485.



68. LANGMUIR, I.: The Adsorption of Gases on Plane Surfaces of Glass, Mica, and Platinum. *J. ACS*, vol. 40, 1918, p. 1361.
69. YOUNG, D. M.; AND CROWALL, A. D.: Physical Adsorption of Gases. Butterworths, London, 1962.
70. BRUNAUER, S.: The Adsorption of Gases and Vapors, Vol. I, Physical Adsorption. Princeton Univ. Press, 1945.
71. LEVINE, O.; AND ZISMAN, W. A.: Physical Properties of Monolayers Adsorbed at the Solid-Air Interface. I: Friction and Wettability of Aliphatic Polar Compounds and Effect of Halogenation. *J. Phys. Chem.*, vol. 61, 1957, p. 1068.
72. KIPLING, J. J.: Adsorption from Solutions of Non-Electrolytes. Ch. 10, Academic Press, 1965.
73. GREENHILL, E. B.: The Adsorption of Long Chain Polar Compounds from Solution on Metal Surfaces. *Trans. Faraday Soc.*, vol. 45, 1949, p. 625.
74. DEFAY, R.; AND HAMMELEN, J. R.: Dynamic Surface Tensions. II: Measurement of Dynamic Surface Tension of Aqueous Solutions by the Falling Meniscus Method. *J. Colloid Sci.*, vol. 14, 1959, p. 401.
75. HOTTA, H.; AND ISEMURA, T.: Studies on the Gas-Liquid Interface by Surface Potential Measurements. III: The Surface Aging of Solutions of Long Chain Dibasic Acids. *Bull. Chem. Soc. Japan*, vol. 25, 1952, p. 101. (Chem. Abs., vol. 47, 1953.)
76. ROWE, C. N.: Some Aspects of the Heat of Adsorption in the Function of a Boundary Lubricant. *ASLE Trans.*, vol. 9, 1966, p. 101.
77. ROWE, C. N.: A Relation Between Adhesive Wear and Heat of Adsorption for the Vapor Lubrication of Graphite. *ASLE Trans.*, vol. 10, 1967, p. 10.
78. RABINOWICZ, E.: Influence of Surface Energy on Friction and Wear Phenomena. *J. Appl. Phys.*, vol. 32, 1961, p. 1410.
79. CAMPBELL, B. D.: Auger Electron Studies of Surfaces: Uranium Dioxide, Uranium, Graphite, 300 Series Stainless Steel, and Niobium. Los Alamos Lab. Univ. Calif., LA-4010, UC 25, March 5, 1969.
80. GROSZEK, A. J.: Preferential Adsorption of Compounds with Long Methylene Chains on Cast Iron, Graphite, Boron Nitride, and Molybdenum Disulfide. *ASLE Trans.*, vol. 9, 1966, p. 67.

# Chemistry in Concentrated-Conjunction Lubrication

R. S. FEIN

Texaco Research Center  
Beacon, New York

"Indifferent" base oils, "lubricity" agents, "EP" agents, oxygen and water all influence wear, load-carrying capacity, and surface fatigue; in some cases the direction of the influence appears predictable. All the above materials react chemically at wearing surfaces yielding products generally containing both organic and inorganic components and oxygen. The reactions apparently result from both thermal and catalytic activation; in addition, mechanical stresses (probably) and exo-electrons (possibly) play roles. The physical state of the reaction products and the probability of their remaining near the wear surfaces depend on their tendency for solubilization by bulk lubricant.

Consideration of the physical mechanisms involved in lubrication, wear, and surface damage indicates the probably important physical properties of wear-surface reaction products. These properties particularly include flow and elastic characteristics (i.e., rheological properties), thermal conduction characteristics, and ability to non-thermally affect substrate metal deformation (i.e., Rehbinder effects). An understanding of chemical effects on wear and surface damage requires development of knowledge about the pertinent physical properties of wear-surface reaction-product films. These properties need to be correlated with system chemistry and operating conditions and with film performance in terms of wear and surface damage. Achievement of these correlations requires development of techniques for micro-scale, in-situ measurement of wear-surface film rheological and thermal conduction properties.

THE CHEMISTRY OF LUBRICATION today is best characterized as an **T**art rather than a science. Lubricants dealing with most problems of concentrated-conjunction\* lubrication can be formulated because the problems represent comparatively minor variations from established practice. The success has been achieved with greater difficulty in dealing with problems representing larger extrapolations from present practice.

---

\* The term "conjunction" rather than contact is used since load-supporting surfaces always approach each other but do not necessarily make contact.

Increasing frequency of difficult-to-solve lubrication problems plus recent developments in chemical purification and analysis techniques apparently are sparking a resurgent interest in understanding the chemistry of lubrication. Recent studies concentrate on the lubrication effects of well-controlled chemical environments for bearing surfaces and on chemical examination of the surfaces and wear debris. These studies contrast with earlier studies of lubricant effects in either very idealized or quite uncontrolled chemical environments. The earlier studies inferred the nature of bearing surface films from (1) comparison of friction and wear with that obtained on preformed films and/or (2) characterization of films formed on model non-bearing surfaces. The earlier studies classified the lubricants by chemical type and additives in terms of their performances and postulated mechanisms of action. Unfortunately, despite contrary indications (refs. 1 to 3), an aura of established fact resulted from repeated quotation of a number of postulates from these earlier studies. Dissipation of this aura began recently as the implications from recent work became known.

This paper will outline current knowledge concerning the chemistry of lubrication in concentrated conjunctions between counterformal bearing surfaces.\* Emphasis will be placed on wear and surface damage to hard steel surfaces operating with hydrocarbon-based liquid lubricants. The emphasis is placed on wear and surface damage rather than on the friction often measured in fundamental studies because these phenomena usually control the useful life of a bearing. Hard steel surfaces and hydrocarbon-based lubricants are emphasized because of their great importance in the concentrated conjunctions of gears and rolling-element bearings. The phenomena discussed also usually occur with other kinds of bearing materials and lubricants. The following sections summarize the effects of lubricant and atmosphere constitution on wear and surface damage, present-day knowledge of the constitution and physical properties of bearing surface films, and the interacting chemical and physical steps apparently involved in wear and surface damage control. This summary is aimed at generalizing and integrating available information and ideas as a basis for conclusions concerning need for further information. No attempt is made to cite all or necessarily the primary pertinent references when discussing particular points.

#### CHEMICAL EFFECTS ON WEAR AND SURFACE DAMAGE

This section covers the effects of lubricant and environmental constitution on wear, load-carrying capacity, and pitting, along with the interaction of these constitutional effects with operating variables. Probably the most characteristic feature of these chemical effects is that no quanti-

---

\* Bearing surface is used in the general sense of any load-supporting surface.



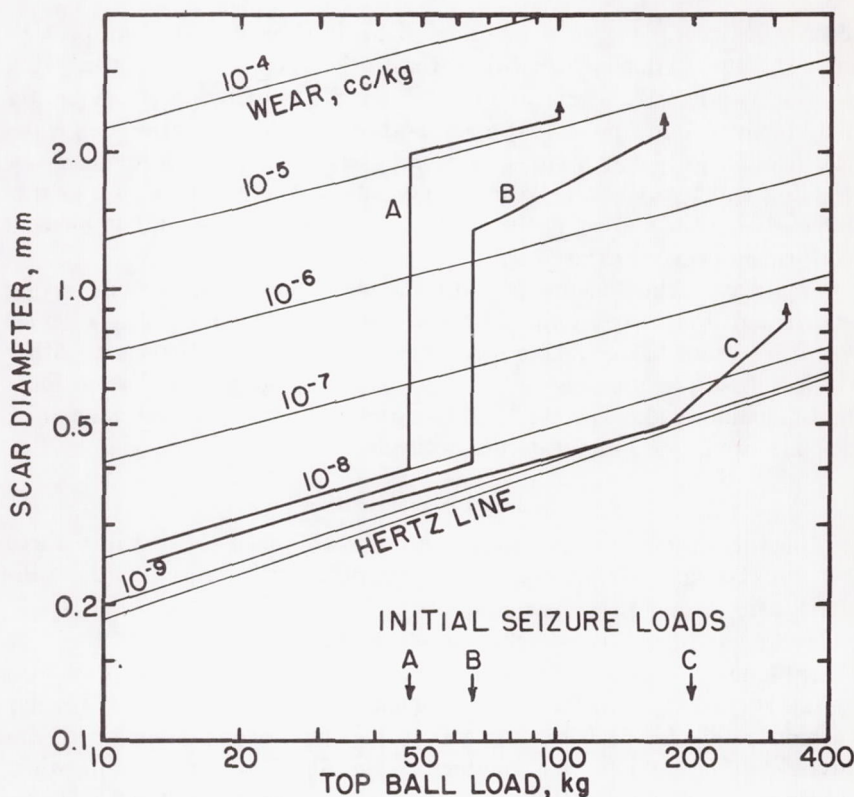


FIGURE 1.—Schematic four-ball machine data. 69 cm/sec, 10 sec, room temperature. Lubricant type: A, indifferent; B, mild EP; C, strong EP. Arrows on curves indicate weld load.

tative generalizations can be made, and reasonably valid qualitative generalizations can be made in only a few cases. The effects illustrated are typical of the kinds that play such an important role in concentrated-conjunction lubrication.

A large fraction of the evidence on the influence of chemistry on wear and surface damage is from four-ball machine (FBM) studies. The major reasons for this preponderance of data from a single machine type seem to be: (1) availability of cheap, reproducible test specimens; (2) small lubricant volume; (3) ease of temperature and atmosphere control; (4) its sensitivity to low wear levels ( $\sim 10^{-8}$  cm<sup>3</sup>/wear scar);\* and (5) the possibility for studying four separate wear and surface damage phenomena.\* Figure 1 shows how the four phenomena are re-

\* Reasons 4 and 5 are not applicable to the "rolling" or "planetary" FBM used to study surface fatigue.

flected in wear level as a function of load. Effective lubrication with smooth low friction occurs below the initial seizure load. Seizure is a friction rise shortly after the start of a run (the friction subsequently may recover to low levels). If a run is stopped during seizure (i.e., while the friction is high), wear scars usually show microscopic evidence of welding and tearing. The wear at loads above the initial seizure load is indicative of the control of "severe" wear. At the weld load, friction and surface damage are catastrophic.

Practical machine elements and other rigs, for various reasons, rarely show more than two of the four wear phenomena. However, judicious use of the four-ball machine often yields useful correlations with other devices (usually unpublished). Thus, results to be discussed from four-ball machine studies, rather than being restricted to four-ball machines, are of more general qualitative significance.

#### Types of Lubricant Materials

The various kinds of lubricant materials are often divided into three general classes: (1) "nonreactive" or "indifferent" materials, (2) "oiliness" agents, and (3) "extreme-pressure" or "EP" agents.

Hydrocarbon and most other base oils fall in the "nonreactive" or "indifferent" class. Usually the most important property of these materials is viscosity, which depends on chemical structure in a fairly regular fashion (refs. 4 and 5). Viscosity, with its dependence on pressure, has been invoked to account for the greater fatigue life of rolling elements when lubricated with a naphthenic oil rather than with a paraffinic oil (refs. 6 and 7).

Recent evidence shows large influences of hydrocarbon type on wear and load-carrying capacity under conditions of pure sliding. Figure 2 shows wear varying more with hydrocarbon type than with viscosity under conditions of negligible frictional heating (ref. 8) and elastohydrodynamic load support on the basis of bulk lubricant properties. Similar results are obtained under conditions of appreciable frictional heating, with unsaturated hydrocarbons (olefins and aromatics) giving lower wear than saturated ones (paraffins and naphthenes) (refs. 9 and 10). The aromatics also provide greater load-carrying capacity than the saturated paraffins and naphthenes as shown in table 1 (ref. 10).

All mineral oils consist of molecules generally containing more than one type of hydrocarbon moiety. Except for olefins which are rare in all but Pennsylvania oils (ref. 4), oils include all the above moieties. In fact, single molecules often will contain paraffinic, naphthenic (i.e. cycloparaffin) and aromatic portions. Classification of oils as paraffinic, naphthenic, or aromatic, indicates the relative concentrations of the moieties. Table 2 illustrates typical distributions of these moieties for highly refined paraffinic distillates. Generally natural sulfur and polar

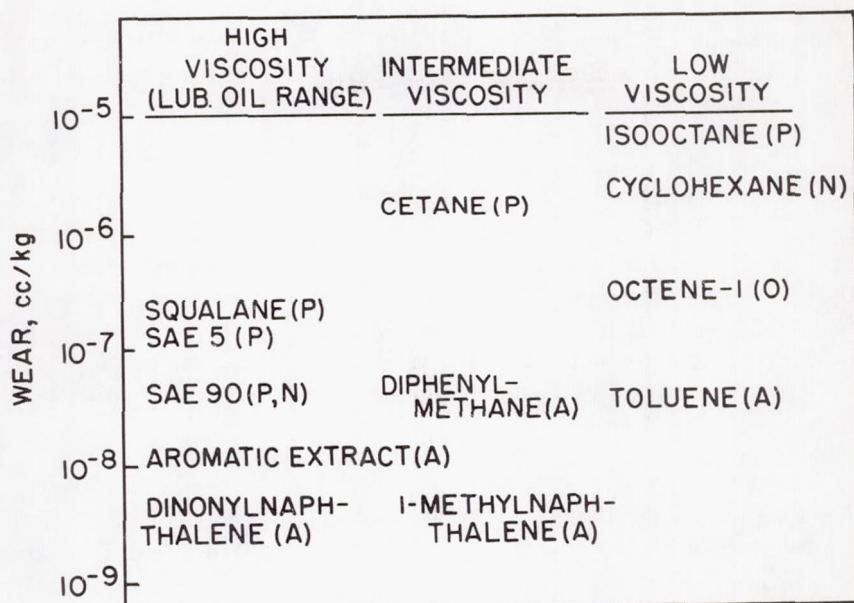


FIGURE 2.—Comparison of wear with hydrocarbons. FBM, 50 kg, 9.2 rpm, 15 to 16 hr, 100° F. Lubricant type: P, paraffinic; N, naphthenic; O, olefinic; A, aromatic.

TABLE 1.—*Effect of Hydrocarbon Type and Atmosphere on Seizure Load\**

Hydrocarbon	Seizure load, kg	
	Air	Argon
Cyclohexane (N)	15	22
n-heptane (P)	37	15
Benzene (A)	52	30
Toluene (methylbenzene) (A)	105	30
Cumene (isopropylbenzene) (A)	157	37

\*FBM, 23 cm/sec. Lubricant type: P, paraffinic; N, naphthenic; A, aromatic (ref. 10).

TABLE 2.—*Typical Paraffinic Neutral Compositions*

Visc., cs		Wt % sulfur	Wt % polar	% Carbons			Av rings per molecule	
100° F	210° F			P	N	A	N	A
33.0	5.44	0.05	0.6	68.2	25.1	6.7	1.50	0.32
73.0	8.39	0.13	0.7	63.7	32.9	3.7	2.27	0.20



compounds in oils reduce wear and surface damage (refs. 11 and 12). For any given series of oils, the concentrations of these materials as well as aromatics often tend to increase with viscosity. This fact probably accounts for much of the often observed decrease in wear and increase in load-carrying with increased viscosity of straight mineral oils.

Additives containing a polar group attached to long chains of methylene groups (i.e.  $(-\text{CH}_2-)_n$  where  $n > \sim 10$ ) are often referred to as "oiliness" or "lubricity" additives; oxygen is the most usual hetero-atom in the commonly used materials. These terms, oiliness or lubricity, reflect the fact that these additives usually lower and smooth the friction and reduce wear under conditions of low sliding and rolling velocities.

Typical lubricity additives are natural fats or oils or fatty acids such

as stearic acid (i.e.  $\text{CH}_3(-\text{CH}_2-)_{16}\overset{\text{O}}{\parallel}\text{C}-\text{OH}$ ). Examples of the friction-reducing properties of fatty acids abound (refs. 13 and 14). The effect on wear is illustrated by adding 0.43 weight percent stearic acid to cetane (a 16-carbon normal paraffin,  $\text{CH}_3(-\text{CH}_2-)_{14}\text{CH}_3$ ) and squalane (an isoparaffin with six  $\text{CH}_3$  groups on a 20-carbon backbone). Volumetric wear is reduced by factors of about 100 and 10, respectively, under the conditions of figure 2. Load-carrying capacity\* measured in terms of the maximum temperature in the load-bearing conjunction (i.e., "critical temperature") is greatly increased by the additive in cetane, but hardly affected in squalane (ref. 15) or a mineral oil (ref. 16). Fatigue life seems to be hardly affected in rolling contact with small slip.\*\* By contrast, in rolling contact with moderate slip, oleic acid greatly increases fatigue life as shown in table 3 (ref. 17). Thus, the oiliness agents can be effective in controlling wear and surface damage, but their effectiveness depends on the base oil and operating conditions. Rather equivocally, their effectiveness seems favored by operating conditions tending to give both thin elastohydrodynamic (EHD) films and low sliding velocities.

The second major class of additives used to control wear and surface damage are the so-called "extreme-pressure" or "EP" additives. These additives usually contain the hetero-atoms phosphorous, sulfur, or chlorine and may additionally or even principally contain oxygen, lead, zinc, boron, selenium, etc. Many authors have pointed out that the effects of these additives are most evident under conditions of high temperature in the load-bearing conjunction; in other words, the effects are most evident under conditions of high sliding velocity, high and concentrated loads, and high bulk temperatures (refs. 18, 19, and 20).

Addition of EP agents to mineral oils, as illustrated in figure 1, often

\* Load-carrying capacity is the combination of operating conditions at which initial seizure occurs.

\*\* Rein, S. W.; and Fein, R. S.: Unpublished, 1960.

TABLE 3.—*Fatigue Life Dependence on Base Oil and Additives\**

Additive	Base Oil	
	A	B
None (I)	87	41
1% oleic acid (L)	113	—
1% dialkylphosphate (EP)	22	203
5% chlorinated wax (EP)	11	58
2.5% sulfurized terpene (EP)	113	—

\*Planetary FBM,  $1.2 \times 10^6$  psi max. Hertz stress, 1580 rpm, 200° F.  $B_{50}$  life with 52100 air-melt steel in minutes. Lubricant type: I, indifferent; L, lubricity; EP, extreme pressure.

TABLE 4.—*EP Effects of Additives in Cetane\**

Wt percent additive	Initial seizure load, kg	Wear, $10^{-6}$ cc/kg		Weld load, kg
		at 50 kg	at 100 kg	
0.43% stearic acid	20	17.	18.	110
None	7	2.6	9.0	110
0.33% dibenzylsulfide	26	6.0	5.2	130
0.15% sulfur	12	0.03	0.02	158
0.375% zinc diisopropyl-dithiophosphate	36	6.7	7.1	120
5% complex borate	92	—	0.47	140
5% lead naphthenate	22	3.8	4.1	170

\* FBM (EP), 69 cm/sec, 77° F, 10 sec.

decreases the wear rate under conditions less severe than required for initial seizure;\* however, the magnitude of the effect and, in some cases, its direction depend on all the system variables including constitution and concentration of other components, and operating variables.\* Further, as illustrated by figure 1 and by wear values at 100 kg in table 4, EP agents most often will control the extent of surface damage under conditions more severe than those producing initial seizure. In many cases, this permits mechanisms such as gears to function satisfactorily during and after operating at conditions above the initial seizure limit.

In addition to wear and surface damage control, EP agents commonly increase the severity of conditions required to produce "seizure" and catastrophic failure by surface damage or "welding." Thus EP agents usually markedly increase the load-carrying, temperature, and/or sliding-

\* Randall, T. H.; and Fein, R. S.: Unpublished, 1961.

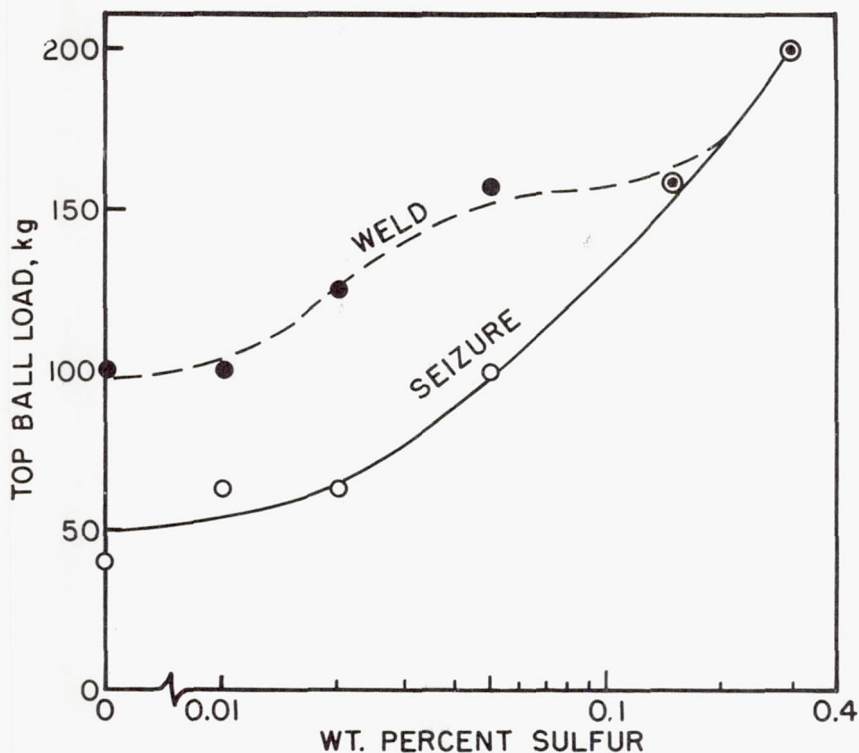


FIGURE 3.—Effect of sulfur concentration in cetane. FBM, 69 cm/sec, room temperature, 10 sec.

velocity capabilities of a lubricant (illustrated by figure 3 and by the large load-carrying abilities of the EP lubricants in table 4.)

Fatigue life may be either increased or decreased by EP agents as illustrated in table 3.

#### Oxygen and Water

Oxygen and water play extremely important but often unappreciated roles in the chemistry of lubrication. These reactive materials are extremely difficult to exclude from a system except via hard vacuum and rigorous outgassing. Their effects on lubrication have been studied fairly extensively in recent years.

Generally for hydrocarbons which are not highly aromatic, wear increases with oxygen availability at conditions less severe than those required for seizure (refs. 3, 8, 21, 22, and 23); figure 4 illustrates this phenomenon (ref. 8). Usually, increasing oxygen concentration increases the severity of conditions required to produce seizure (i.e., oxygen in-



creases load-carrying capacity) (refs. 8, 21, and 22), illustrated for hydrocarbon lubricants by table 1.

Generally similar trends of seizure limits with oxygen availability are found with fatty acids (ref. 10) and sulfur-containing EP agents (ref. 24). Other EP agents also seem to respond to oxygen, but not consistently (ref. 22). Increasing oxygen concentrations seem to have a favorable effect on the amount of wear and extent of surface damage under conditions more severe than those for initial seizure (ref. 10), and, at least with tricresyl phosphate, under conditions less severe than initial seizure (ref. 12). The influence of oxygen on fatigue life does not appear to have been established.

The oxygen effects discussed in the foregoing paragraph occur in the absence of bulk lubricant oxidation. The latter, of course, is promoted by increasing oxygen concentration and temperature. The occurrence of appreciable bulk lubricant oxidation in some cases degrades lubrication; for example, figure 5 (ref. 25) shows an abrupt increase in wear rate (below the initial seizure limit) when the temperature is sufficiently high to give bulk oil oxidation (indicated by formation of carbonyl groups absorbing at  $1710\text{ cm}^{-1}$ ). In another study, fatigue life of a bearing lubricated with a naphthenic oil increased when oil oxidation was inhibited (ref. 26).

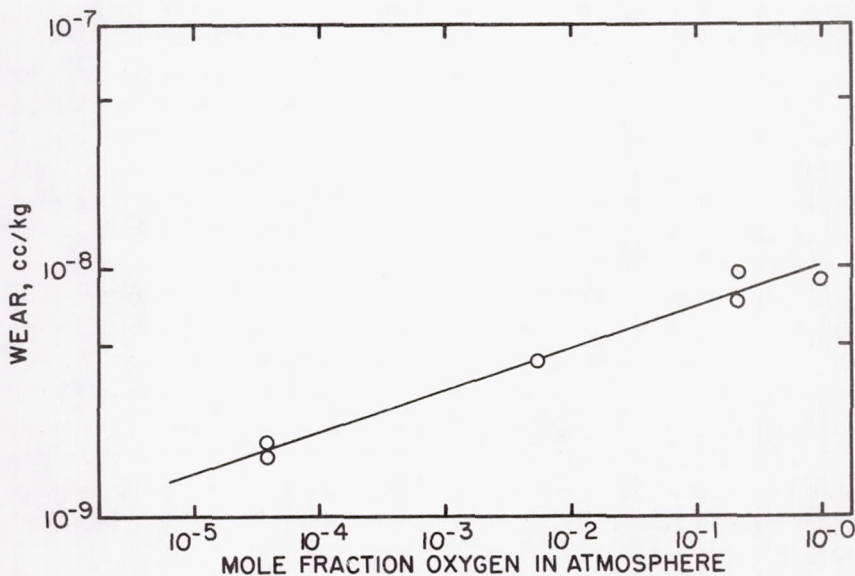


FIGURE 4.—Effect of oxygen on wear. FBM, 9.2 rpm, 5 kg, 10m,  $25^{\circ}\text{C}$ .  
Lubricant: squalane.

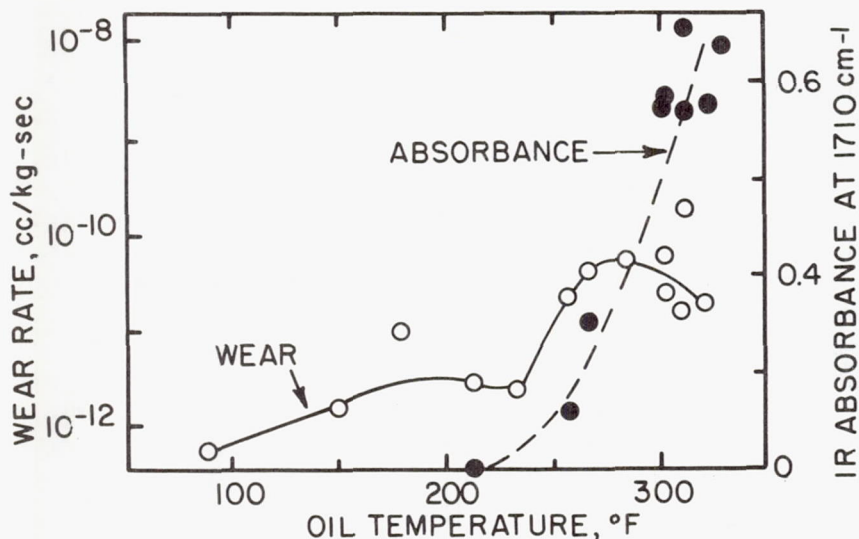


FIGURE 5.—Wear and used oil changes. FBM, 23 cm/sec, 27 kg, 2 hr. Lubricant: 0.43-percent stearic acid in cetane (ref. 25).

Effects of water on wear and surface damage (other than fatigue) have not been studied as thoroughly as those of oxygen. For the most part, however, the effects that have been studied seem similar. Thus, for example, oxygen and water both tend to raise the scuff-limited load with aromatic lubricants at high sliding speeds in a ball-on-cylinder machine (ref. 27). Also, water as well as oxygen can produce effective lubrication with fatty acid lubricants (ref. 28).

There are other effects of moisture, however, which seem distinct from those of oxygen. Figure 6 shows the effect of humidity on wear below seizure with an EP agent in two different carriers. Effects of the humidity on wear were not evident in runs shorter (e.g., 10 minutes) than those shown in the figure. Further, the important humidity seemed to be that of the ambient atmosphere at the time the test specimens were cleaned,\* rather than that used to saturate the oil during testing. Apparently, conditioning of the bearing surfaces with moisture can have a long-range effect on wear.

Another effect of moisture different from that of oxygen is the effect on rolling contact fatigue. It is well established that increasing moisture levels decrease fatigue life markedly (refs. 7 and 29). Table 5 shows this effect.

\* Cleaned with stoddard solvent, dried with purified isooctane, and immediately covered with test lubricant after cleaning.

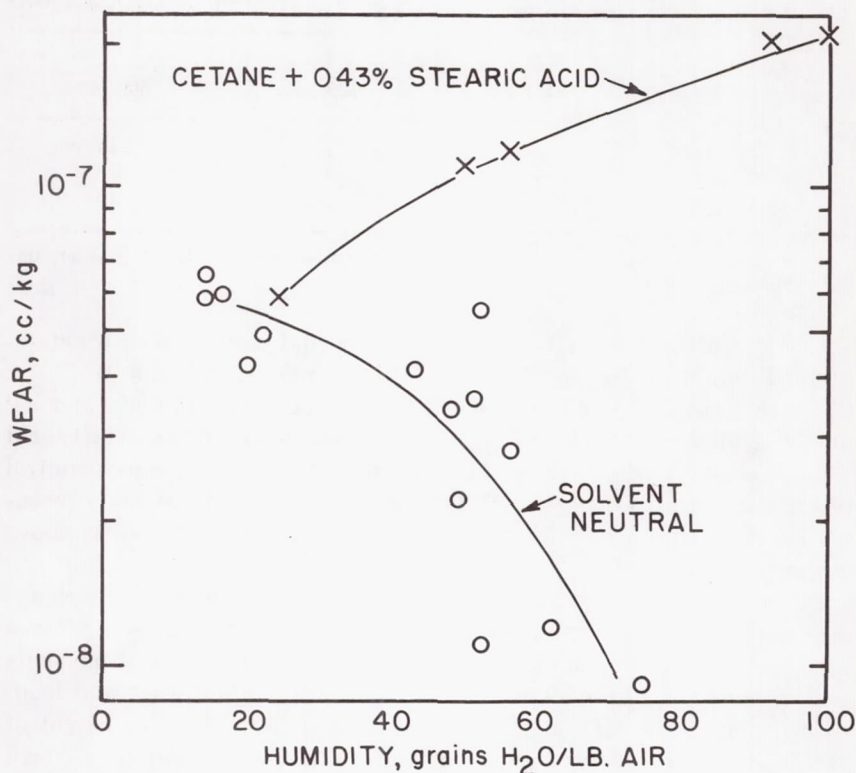


FIGURE 6.—Effect of humidity on wear. FBM, 27 kg, 23 cm/sec, 210° F, 2 hr. Additive: zinc diisopropyldithiophosphate (0.05% Zn).

#### Interaction of Chemical Components

A well-known feature of lubrication chemistry is that any chemical component may compete or interact with any other component. Further, the extent to which the interactions are important depends upon operating conditions and bearing materials. For the most part, the interactions seem to be characteristic of specific systems rather than of general phenomena. There does, however, seem to be one general principle; two materials which tend to adsorb and/or react on bearing surfaces will often compete with, rather than complement, each other.

An example of this competition is the influence of refining on the additive response of mineral oils. Generally the response to lubricity and EP additives (also anti-oxidants and anti-corrosion additives) improves with the degree of refining. Most refining processes reduce concentrations of aromatic hydrocarbons and molecules containing the hetero-atoms, sulfur, oxygen, and nitrogen. Apparently the natural antiwear and EP



TABLE 5.—*Effect of Lubricant Moisture on Fatigue Life of Tetraester Oils*

Additive package	<i>B</i> <sub>50</sub> life in hours at approx moisture content*		
	50 ppm	225 ppm	500 ppm
A	130	120	35
B	—	110	25
C	68	52	48

\* 204 bearings, 10 000 rpm, 360 000 psi max. stress, 400° F. (R. C. Muller, unpublished data, 1968).

activity of the naturally occurring aromatics and molecules with heteroatoms interferes with the effect of additives (refs. 12, 30, and 31).

Another example of competition is the interaction of lubricity and EP additives illustrated in figure 7 (ref. 11). The low concentration of fatty acid reduces wear prior to seizure, but does not affect the EP-agent control of surface damage above seizure. On the other hand, a large acid concentration interferes with the EP agent's scuff-preventing ability above seizure.

Other additives may also interact in addition to base-oil, lubricity-agent, and EP-agent interactions. Most notable among these additives are the amphipathic polar-group containing materials used as dispersants and detergents. These often will degrade both the antiwear and load-carrying action of lubricity and EP additives (ref. 11). The extent of the degradation depends greatly on the particular combination as shown in table 6.\*

#### CONSTITUTION OF WEAR PRODUCTS

As discussed earlier, a major cause of the recently renewed interest in lubrication chemistry has been the development of more powerful analytical techniques that permit characterization of the minute quantities of reaction products formed in effectively lubricated systems. In this section, results of characterizations will be summarized. The reaction products to be discussed are found as films on the bearing surfaces or as wear debris sludges that may remain suspended or be deposited on various surfaces or in filters. In the case of the sludges, it is assumed that the sludge material is representative of wear-surface film material which has been removed by the wear.

In the chemical characterization of wear products, extreme care must be exercised not to contaminate or chemically change the product inadvertently (ref. 32). Since quantities of wear product are commonly in the 10 to 1000 microgram range, very small amounts of contamination

\* Kreuz, K. L.; and Fein, R. S.: Unpublished, 1957.

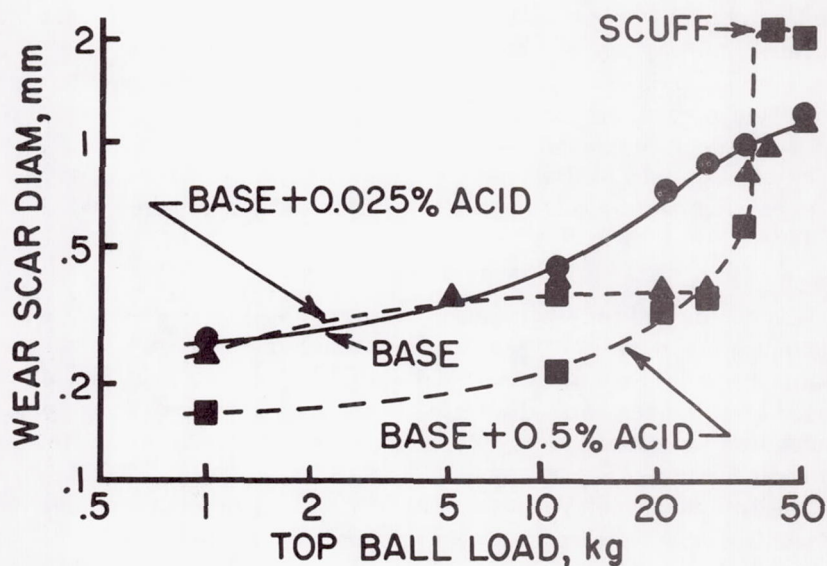


FIGURE 7.—Fatty acid-EP interactions. FBM, 69 cm/sec, 210° F, 10 min. Base: cetane plus 0.375-percent zinc diisopropylthiophosphate. Acid: behenic acid (ref. 11).

from dust, finger prints, or impure solvents must be avoided. Products generally are physical dispersions of inorganic materials in lower oxidation states (e.g.,  $\alpha$ -iron or  $\text{Fe}_3\text{O}_4$ ), metalloorganic complexes (e.g., iron soaps) and/or organics (e.g., base oil, additives, oxidation products). Contamination can be avoided with sufficient caution; however, chemical change avoidance depends on the nature of the material. Deoiling and transfer of the material by solvent treatment is usually necessary for characterization of organic and metalloorganic wear products; however,

TABLE 6.—*Detergent-EP Additive Interaction\**

Basic sulfonate (8.8meq/100g)	Wear, $10^{-9}$ cc/kg	
	No EP	Zinc diisopropylthiophosphate (0.07% zinc)
None	308	7 to 33**
Ca	133	15
Ba	137	24
K	104	75

\* FBM, 27 kg, 23 cm/sec, 210° F, 2 hr. Light solvent neutral oil.

\*\* Sensitive to humidity.

dealing of accompanying inorganic products (e.g.,  $\alpha$ -iron and its lower oxides) permits their oxidation (e.g., to  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub>). Further, the better solvents (particularly the organic bases) may decompose or otherwise react with metalloorganic complexes (e.g., "ferric stearate" may be progressively decomposed to oxide and stearic acid).<sup>\*</sup> Each wear product must be examined carefully to avoid fallacious characterization.

Physical characterization of the wear products must also be approached with caution because solvents often will change their colloidal nature.

#### Hydrocarbon Wear Products

Wear products from hydrocarbon lubricants tend to be predominantly organic when wear is low and predominantly inorganic when wear is high (ref. 8). The organic material consists of saturated hydrocarbon (even from the unsaturated olefinic and aromatic lubricants!), carbonyl-containing acids and esters, and usually carboxylate soap moieties in complex mixtures (refs. 8, 20, and 25).

The amount of oxygen incorporated in the organic material and the amount of soap increase with available oxygen and with the amount of wear when operating in the effective lubrication regime (ref. 33). The wear product is polymeric in nature as evidenced by its refractory nature and (from a low molecular weight hydrocarbon) actual molecular weight measurements (ref. 33). In fact, these rather nondescript wear products are commonly termed "friction polymers" (ref. 8).

The constitution of the wear products' inorganic portion seems to be more sensitive to surface environment than the organic portion. Thus, with increasing oxygen availability, the inorganic wear product changes from iron and/or an iron "carbide," "FeC," to iron oxide.<sup>\*\*</sup> At low oxygen concentration, low-molecular-weight saturated hydrocarbons (i.e., paraffins and naphthenes) and higher-molecular-weight, straight-chain saturated hydrocarbons yield predominantly iron with some FeC and lower oxides; at high oxygen concentrations (e.g., with air atmospheres), they yield  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub> with some lower oxides. Higher-molecular-weight, branched-chain paraffins yield more FeC at low oxygen concentrations and often predominantly Fe<sub>3</sub>O<sub>4</sub>, rather than  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub>, at high oxygen concentrations.<sup>\*\*\*</sup> Aromatic hydrocarbons seem to yield almost completely FeC at low oxygen and Fe<sub>3</sub>O<sub>4</sub> at high oxygen, with mixtures at intermediate oxygen levels (ref. 8). Curiously, FeC is a rather poorly characterized material first found on iron oxide catalysts used in synthesizing hydrocarbons from carbon monoxide and hydrogen. Attempts to com-

---

<sup>\*</sup> Private communication with K. L. Kreuz.

<sup>\*\*</sup> FeC is identified by its distinctive X-ray diffraction pattern (ref. 34).

<sup>\*\*\*</sup> X-ray diffraction patterns from wear products do not distinguish  $\gamma$ -Fe<sub>2</sub>O<sub>3</sub> from Fe<sub>3</sub>O<sub>4</sub>.



pletely separate FeC from friction polymer suggest that FeC may be an inorganic complex of iron and an unsaturated hydrocarbon ligand. \*

#### Lubricity Agent Wear Products

Reaction products formed by lubricity agents seem to have been studied extensively only for fatty acids and, to a limited extent, carboxylic esters. In these cases, carboxylate soaps are formed when oxygen and/or water is present (ref. 11). The soaps formed on steel surfaces have often been found to contain ferrous iron despite the usual ease of iron oxidation (ref. 33). When iron is found in the ferric state it is not clear whether the wear product originally existed in this form or oxidized in handling. To this author's knowledge, the stoichiometry of the iron soaps in wear products has not been thoroughly studied. Thus, there seems to be no information on whether the soaps are neutral  $[(R-CO_2^-)_2Fe^{++}]$ , basic  $[(R-CO_2^-)Fe^{++}(OH^-)]$ , or overbased  $[(R-CO_2^-)_2Fe^{++} \cdot xFeO]$ . Certainly, basic metal soaps may be made outside wearing systems (ref. 35), and attempts to prepare ferric stearate produced material covering a wide range of compositions containing excess iron.\*\* The iron soaps formed synthetically, and those found on wear surfaces seem to be insoluble in common solvents, but often dispersible in good solvents.

#### EP Agent Reaction Products

Recently obtained facts have been exploding fictions concerning the constitution of wear products formed from EP agents. In three cases where wear-surface film constitution has been thoroughly studied, film compositions have been found grossly different from those commonly supposed. First, careful studies show that solid wear-surface films formed from sulfur EP agents contain principally iron oxide and organic material with a comparatively small fraction of sulfur as the sulfide ion (refs. 34 and 36); this contrasts with the fiction that iron sulfide films are formed.

Second, the films formed from phosphate esters seem to be organo-metallic phosphates (refs. 37 and 38), with some iron phosphate identifiable when the lubricant is 100 percent additive (ref. 39). No iron phosphide could be found in the films despite the fact that it should be there according to the commonly quoted mechanism of phosphate action.

The last example concerns the wear surface films from borate esters. These less familiar EP agents were postulated to decompose yielding simple boric oxide polymers on wear surfaces (ref. 40). In fact, they were found to produce a film with about as many atoms of iron and carbon as of boron (ref. 41). Apparently the iron and boron and con-

\* Private communication with M. Dundy.

\*\* Private communication with K. L. Kreuz.

siderable oxygen were dispersed in an organic carbon-based matrix, since they could be extracted with dilute acid without disintegration of the film's physical structure. The remaining carbon-based matrix showed infrared bands of carbonyl and hydroxyl groups but not well-defined carbon-hydrogen bands.

The elemental composition found in these three examples where films have been sufficiently analyzed are summarized in table 7. In all cases there is roughly as much organic as inorganic material in the film. Secondly, oxygen and iron are important film constituents. Thirdly, the EP element, although always present in the film, is not necessarily a major component.

Figure 8 shows a transmission electric micrograph of a typical EP film after removal from a surface operating under conditions of high sliding velocity and high unit load (refs. 34 and 41). This typical EP film shows the heterogeneity of the films and a "lay" probably associated with the original bearing surface lay. The films are typically of the order of  $0.1\text{ }\mu\text{m}$  (i.e.  $1000\text{ \AA}$ ) thick as evidenced by their transparency under electron beams. Similar orders of magnitude film thicknesses have been estimated from EP-element concentrations (refs. 42 and 43).

The concentration of EP elements on surfaces has been found generally to range between 1 and  $100\text{ }\mu\text{g}/\text{cm}^2$  (refs. 31, 34, and 42 to 45). Further, with any given lubricant, concentrations generally increase with operating conditions which increase surface temperatures; thus, EP element concentrations increase with Hertz pressure, sliding velocity, and bulk lubricant temperature. Further, the concentrations seem to build to a steady-state level under steady operating conditions (ref. 42). Although not proven, it is often assumed that the highest concentrations of film ma-

TABLE 7.—*EP Film Constitution*  
Approximate Atomic Ratios in EP Films

EP type	Sulfur (ref. 34)	Borate (ref. 41)
Iron	1.00	1.00
Oxygen	1.2	4.7 (by difference)
Carbon	0.5	1.0
EP element	0.02 (sulfur)	0.82 (boron)
Hydrogen	0.8	Not determined

"Organic" to "Inorganic" Phosphate in Effective Phosphate Ester Film (ref. 37).

Surface Material:

Chrome Plate	3.1
Hardenable Iron	1.3
Chilled Cast Iron	0.3

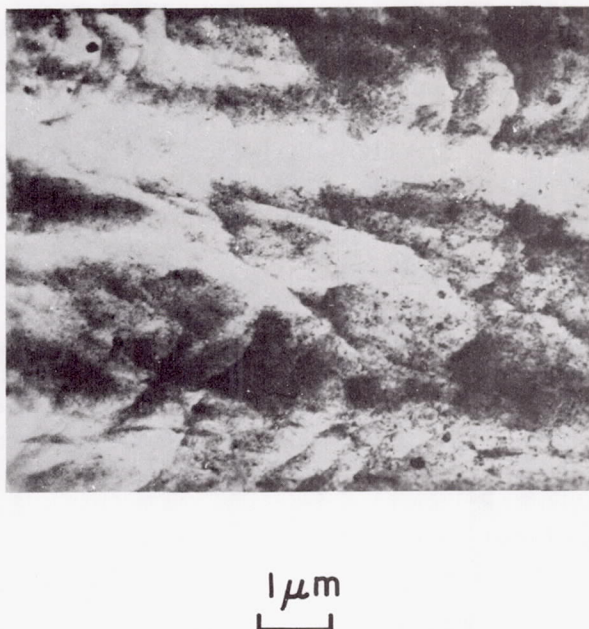


FIGURE 8.—Transmission electron micrograph of EP film.

terial are on the highly-loaded asperities. The high temperature on these asperities is assumed to produce chemical reactions accounting for the observed nonuniform distribution of EP elements on bearing surfaces (ref. 42).

Three additional features of the reaction products from EP agents warrant attention. First, with sulfur EP agents, sulfur (ref. 43) and oxygen (ref. 34) appear to penetrate the bearing metal to depths of the order of several micrometers; confirmation of this effect is needed. Second, with EP agents, trace elements from the bearing metal may concentrate in the EP film; thus, with sulfur EP agents a much higher silicon-to-iron ratio is found in the film and wear debris than in the steel (ref. 34). Third, with EP agents, inorganic wear products may be produced which are not otherwise generally found in wear-surface films. Thus, with sulfur EP agents, crystalline diffraction patterns indicated an unknown spinel (ref. 34) and  $\text{SiO}_2$  (refs. 34 and 46). Additionally sulfur (at least in cetane) changes the dominant inorganic wear product from  $\alpha\text{-Fe}_2\text{O}_3$  to  $\text{Fe}_3\text{O}_4$  (ref. 46). With a borate ester in cetane,  $\text{Fe}_3\text{C}$  was frequently observed in four-ball machine wear debris after operation in ambient air; this carbide is otherwise only rarely found in wear products.\*

\* Fein, R. S.; Kreuz, K. L., and Dundy, M.: Unpublished, 1963.



#### **Influence of Other Lubricant Materials and Interactions on Products**

Studies of elemental concentrations in bearing-surface product films and wear debris show profound effects of all constituents. Thus detectable elements from detergents and dispersants are found on wear surfaces (ref. 47). Detergent or dispersant additives can decrease EP element concentration (ref. 48). Model nonhydrocarbon base oil constituents and aromatic hydrocarbons reduce the concentration of EP agent hetero-elements (S, P, and Cl, ref. 31). Of the model non-hydrocarbon constituents, a fatty acid lubricity agent was most effective. Chlorinated wax added to a sulfur EP agent oil can suppress sulfur concentration on the surfaces (refs. 42 and 43). Depending on operating conditions an alkyl phosphate may or may not suppress sulfur concentration (ref. 43).

The interactions of additives are so complex that only a few generalizations are possible. Basically, as expected, concentrations of elements from an additive in wear products usually increase with additive concentration in the bulk oil. Further, amphipaths such as fatty acids and some detergents in sufficient concentration usually decrease the concentration of EP elements in the wear surface films.

#### **CHEMICAL MECHANISMS INVOLVED**

The preceding sections summarized available information on the effects of ambient chemistry on wear and surface damage and on the nature of reaction products formed at wearing surfaces. The present and following sections will attempt to delineate the chemical and physical steps that are probably involved and to rationalize some of the lubrication effects in terms of chemical and physical phenomena. The rationalizations must be to a large extent speculative since most experimental studies have not examined these phenomena directly or considered the interactions of phenomena.

##### **Reactions at Clean Metal Surfaces**

The fact that steel surfaces oxidize readily in the presence of oxygen or water is generally recognized; however, it does not appear to be widely recognized that atomically clean metals react with almost everything they contact. Adhesion studies, for example, show that even mutually insoluble metals will adhere strongly if they are clean (ref. 49); this adhesion or welding is indicative of chemical (i.e., metallic) bond formation and appears to be the chemical reaction involved in the scuffing process.

Reactions more pertinent to effective lubrication also occur on atomically clean metals. Very few of these reactions have been studied with iron or steel; however, a sufficient number of metals have been studied to predict quite safely similar reactions on steel (ref. 50). Thus, hydrocarbons form carbon-metal bonds by chemisorbing on oxide-free metal,

and break carbon-hydrogen and carbon-carbon bonds in the process. Hydrogen atoms presumably can remain adsorbed on the surface or dissolve in the metal. Hydrocarbon species of carbon number lower and higher (i.e., polymers) are formed, and in some instances, completely dehydrogenated "surface carbides" may form. Increasing unsaturation of hydrocarbons facilitates their chemisorption. Oxidation of the metal surface inhibits the chemisorption. On the other hand, oxygen reacts with the labile carbon-metal bonds of chemisorbed hydrocarbon to yield oxygen containing groups. These groups are similar to those found in friction polymer and, in fact, catalytic action by clean surfaces seems to explain friction polymer formation (refs. 8 and 33).

Studies of adsorption of carboxylic acids on several metal catalysts show strong chemisorption (ref. 50). In the absence of oxygen or water, the carboxyl group of the acid appears to be covalently bonded to the metal. Addition of either oxygen or water converts the covalently bonded soap to the conventional ionically bonded soap. From this it appears that the necessity for oxygen and/or water to achieve optimum lubrication with fatty acids (ref. 50) must be associated with the nature of the chemisorption bond.

Studies with organic sulfides and disulfides on nickel catalysts show that these molecules can sulfide the surfaces at low temperatures and can leave chemisorbed hydrocarbon moieties on the surface (ref. 50).

#### Other Methods of Initiating Reactions of Surface With Lubricant

The preceding section indicated that catalysis by clean metal surfaces could account for friction polymer and soaps such as those found on wear surfaces, and presumably could initiate other reactions. However, other initiation methods exist and will be discussed below.

Lubricant materials which are acidic react with iron oxides to yield salts; thus fatty acids, acid phosphates, etc., react spontaneously to produce products such as carboxylate and phosphate soaps. Esters generally are subject to hydrolysis reactions yielding acid products (ref. 32) which can then react with the surface. Presumably this accounts for the positive correlation between hydrolytic stability and EP activity of neutral organic phosphates (e.g., tricresyl phosphate, ref. 37) and neutral organic borates.\*

Liquid lubricant materials react thermally yielding very reactive intermediates such as free radicals which then produce a myriad of products both in the liquid and at the surface. Oxidation is the most common of these thermally initiated reactions; however, an inverse correlation of the thermal stability of zinc dithiophosphates with load-carrying ability leads to the conclusion that thermally produced reaction products provide the

---

\* Dadura, J. G.: Unpublished, 1961.



effective lubrication (ref. 51). In fact, thermally initiated reactions are commonly considered (at least implicitly) to be the main reactions of importance in EP film formation; they certainly appear to explain the increase in EP film thickness with operating conditions leading to increased temperature in the load-bearing conjunction.

Two other proposed reaction-initiating mechanisms are more speculative. One suggests that the mechanical shear stress generates free radicals by tearing molecules apart (ref. 52). Such effects have been observed during high polymer tensile tests (ref. 53). Similar and even higher shear stresses are obtained on lubricant films in concentrated conjunctions (ref. 54). Thus, a friction coefficient of 0.05 at a maximum Hertz stress of 250,000 psi corresponds to a shear stress of 12,500 psi =  $8.6 \times 10^8 \text{ d/cm}^2$ ; this stress is about 10 percent of the glass-like shear moduli of organics and well above the 0.1 to 1 percent fracture strength shown generally by materials (ref. 56). Much more gentle shear stresses have been shown to decompose chemisorbed zinc diethyldithiophosphate (ref. 57). This speculative mechanism therefore seems probable but needs explicit study.

A second proposed reaction-initiating mechanism is electron bombardment of the lubricant by exo-electrons emitted from a bearing surface (ref. 58). Adsorption of hydrocarbons and reduction of friction are enhanced by surfaces showing exo-emission (ref. 59); however, a cause-effect relationship has not been demonstrated. Thus this proposed mechanism also needs study.

#### Other Chemical Reactions

Other chemical, and in some cases autocatalytic, reactions than those mentioned have either been found important or can conceivably be important to concentrated-conjunction lubrication. Foremost among those shown to be important are catalytic reactions involving wear products. Wear products (apparently ferrous soaps) formed at wearing surfaces from oxidation products can catalyze bulk lubricant oxidation and modify the nature of the oxidation products (ref. 25); the oxidation products in turn lead to increased wear (ref. 25). Some oxidation inhibitors may inhibit oxidation by complexing with the wear-product oxidation catalysts (ref. 60).

Thermal decomposition of some carboxylate soaps has been shown to yield metal which then catalyzes decomposition (ref. 61). Metals and metal oxides are reputed to catalyze decomposition of organophosphorous compounds (ref. 62), and  $\alpha\text{-Fe}_2\text{O}_3$  nuclei are reported to catalyze the solid state oxidation of  $\text{Fe}_3\text{O}_4$  to  $\alpha\text{-Fe}_2\text{O}_3$  (ref. 63).

The extremely high pressures in combination with high shear stress between hard surfaces in concentrated conjunctions might be responsible for some of the unidentifiable crystalline wear products. High-pressure



solid-state studies show that shear induced transformations can produce new crystal structures which are metastable at normal pressures (ref. 64). At least in some cases, high-pressure reactions appear to be essentially instantaneous and unusual because the high stresses force interpenetration of electron clouds surrounding atoms and molecules (ref. 64). The extent to which such reactions may be important to lubrication chemistry remains to be investigated.

An important aspect of the chemical steps in lubrication is the extent to which reaction products and wear particles are solubilized or dispersed by the bulk lubricant. Accumulation of viscous liquid-like polymers or suitable viscous dispersions of solid particles at the bearing surface obviously could provide additional hydrodynamic lift. In fact, studies with pure paraffin and naphthene hydrocarbon lubricants show that friction polymers formed at the beginning of a run tend to be soluble in the parent hydrocarbon (refs. 8 and 33). In this case and some others, reduction of lubricant volume near the wear surfaces greatly reduces wear.\* This means of assuring a sufficient supply of reaction product by restricting available solubilizing parent lubricant probably is one of the reasons for the practical success of mist lubrication.

Another example of the undesirable effects of solubilization is probably the effect of amphipaths such as fatty acids, detergents, and dispersants on the load-carrying ability of EP agents (ref. 11). Here presumably the amphipaths tend to remove the EP film from the bearing surface.

Solubilization also may have desirable effects. As previously mentioned, friction polymer from aromatic hydrocarbons appears to contain inorganic particles which seem to be dispersed in the polymer and probably serve to increase its viscosity (ref. 8). The ineffective lubrication of flooded benzene at low ambient oxygen concentration may be due to the lack of sufficient polar oxygen-containing groups to adsorb and disperse the inorganic (principally FeC) particles. In some cases, the influence of carboxylic acids and esters on wear has been attributed to their ability to disperse oxide wear particles (refs. 65 to 67).

In addition to their probable particle-dispersing ability, fatty acid lubricity agents probably function at least partially because their iron soaps are sparingly soluble or dispersible in hydrocarbons. Thus, soap worn from the surface will tend to remain in the vicinity of the surface and with solubilized bulk lubricant may form a grease-like surface film (ref. 68).

#### PHYSICAL MECHANISMS OF LUBRICATION, WEAR AND SURFACE DAMAGE

The chemical steps discussed in the preceding section must affect wear and surface damage by physical means; therefore, the physical

\* Fein, R. S.; and Kreuz, K. L.: Unpublished, 1967.

phenomena involved in effective lubrication, wear, and surface damage are examined in this section. This examination leads to some desirable physical properties for lubricant films, wear surfaces, and wear debris and, in the next section, to some estimates of the extent to which these properties are provided by chemical and physical constitutions observed in bearing systems.

Effective lubrication (i.e., low wear and surface damage and, usually, low friction) results when sufficient low flow resistance lubricant separates bearing surfaces. This low flow resistance lubricant is often considered as fluid- and/or plastic-like; it may be bulk lubricant, chemical reaction products and/or a deposited film (a preformed solid lubricant film). The bearing surfaces are solid-like; they may consist of the bearing metal, immobile chemical reaction products, and/or deposited films. Also, any solid-like wear debris or adventitious dirt particles which are large enough to support some of the load between confining surfaces may be considered as a bearing surface.

The above descriptions in a general sense reflect the literature on both hydrodynamic and boundary lubrication. For hydrodynamic lubrication (EHD in concentrated conjunctions), surface separation is enhanced by high lubricant viscosity at the conjunction entrance conditions, large relative radii of curvature of the confining solid surfaces, and high entrainment velocity (ref. 69). For boundary lubrication, desirable lubricant films are usually considered to be those which become plastic-like at low stress (i.e., low shear stress), are coherent (i.e., flow without fracture) and adhere to the solid surface (ref. 70). Adherence is qualitatively analogous to viscosity in that it reflects the tendency for a lubricant film to be carried into a conjunction between solid surfaces rather than to be forced away; however, to the author's knowledge, the constitutive equation coefficient(s) describing this property has not been defined for the plastic-like film. In the absence of other information, the following assumes that the pertinent constitutive coefficient(s) for adherent plastic-like lubricants depends on temperatures, pressure, etc, in a similar manner to viscosity for liquid-like lubricants.

In general, boundary lubricants are considered to be thin surface films. Hence, hydrodynamic lubrication in the presence of surface films must be considered. Simple analysis, such as Appendix A, indicates that Newtonian liquid-like surface films should act as (1) extensions of the solid substrate if they are sufficiently viscous compared to bulk oil viscosity and/or thin compared to bulk oil film thickness, or (2) hydrodynamic flowing liquid lubricants if the bulk-oil film thickness is negligibly small. In either case, the solid surface separation is increased by a viscous boundary film. In an intermediate bulk-oil film thickness region, both the boundary film and bulk oil flow should be appreciable and might be ex-



pected to interact in a complex manner. In this region the stability of lubricant layers may be critical.

A flowing lubricant film, in addition to being carried into a conjunction, should minimize the stresses on the solid surfaces. Minimizing shear stresses reduces energy dissipation, and thereby the temperature increases, tending to make the surfaces less solid-like. Further, minimizing shear stresses minimizes the combined shear and normal stresses tending to plastically deform the surfaces (ref. 71). Under isothermal conditions (i.e., low sliding velocities), shear stress should depend on the flowing lubricant's viscosity at the hydrostatic pressures in the conjunction. Under nonisothermal conditions (i.e., modest to high sliding velocities), shear stress will be reduced by use of a lubricant which has low thermal conductivity and a large temperature dependence of viscosity (refs. 72 and 73). Reduction in bearing-surface thermal conductivity and in pressures between bearing surfaces also should reduce shear stress; the reduced conductivity raises flowing film temperatures and the low pressures produce smaller increases in film viscosity.

Normal stresses on bearing surfaces should be minimized to reduce their contribution to the combined stresses tending to plastically deform the surfaces. Further, the normal stresses largely determine the hydrostatic pressures in the conjunction and thereby the shear resistance of the flowing viscous lubricant. Considering macro-conjunctions between bearing surfaces, the normal stresses may be reduced by decreasing loads, decreasing elastic modulus, and increasing relative radii of bearing surface curvature (ref. 69). The elastic modulus and curvature effects also hold true for micro-conjunctions between asperities or between third bodies and confining bearing surfaces (ref. 74). In addition, increasing the mean macro bearing-surface separation (e.g., by increased EHD film thickness) will reduce loads and normal stresses at the micro-conjunctions.

Wear seems to occur by the three basic mechanisms of corrosion, fatigue, and plowing (refs. 74 and 75). Corrosive wear results when flowing lubricant film material containing chemically reacted bearing metal is forced aside at a load-bearing conjunction. Fatigue occurs when solid bearing-surface material fractures from the substrate as a result of repeated stress; under effective lubrication conditions, stress transmittal to the solid surfaces presumably occurs via an intervening flowing lubricant film (ref. 74). Plowing wear results when strong (and presumably small curvature) asperities or third bodies cause the production of loose wear particles by "cutting" the bearing surface. As has been pointed out, more than one of these three mechanisms usually functions in any particular wearing system (ref. 68).

Surface damage refers to the degradation of surface contours by excessive wear or plastic flow. In terms of macro-conjunctions, surface damage is most often evident as loss of dimensional tolerances, spalling, or



scuffing. In terms of micro-conjunctions, surface damage is evident as scratching (from plowing) and micro-pitting and sometimes material transfer which affects the micro-geometry of the surface. Scuffing and material transfer involve adhesion of two solids. As discussed recently, adhesion seems to require sufficient plastic deformation of two opposing surfaces to give mating atomically clean areas that are large compared to atomic dimensions (ref. 74).

All of the various wear and surface-damage mechanisms should be ameliorated by factors producing increased flowing lubricant film thickness and decreased stresses (ref. 74).<sup>\*</sup> Further, the likelihood of adhesion should be reduced by coherent deformation of any films interposed between surfaces, since this increases the amount of plastic deformation required to expose substrate metal. Scuffing should be reduced by limiting the extent of plastic deformation.

#### CHEMICAL CONTROL OF WEAR AND SURFACE DAMAGE

The preceding section summarized the physical characteristics of lubricant films, bearing surfaces, and wear debris which apparently can control wear and surface damage. This section discusses the effects of estimated physical properties associated with various lubricant and wear product constitutions. The physical properties are estimated from general knowledge of materials since pertinent physical properties of wear products have not been measured (see tables 8 and 9). In discussing the effects of physical properties on wear and surface-damage phenomena, the size scale of the physical phenomena plays an important role. Hence figure 9 summarizes available information on the sizes involved in effective lubrication at concentrated conjunctions; under ineffective lubrication conditions, gross surface damage of scuffing or pitting usually occurs on a scale larger than  $10^{-2}$  cm.

One of the means by which lubricants reduce wear and surface damage is by increasing the separation of the solid surfaces.<sup>\*\*</sup> As previously indicated, a surface film more viscous than the bulk lubricant can serve to increase surface separation. Considerable evidence seems to demonstrate the adsorption of viscous boundary films up to about  $0.1\ \mu\text{m}$  thick from oils containing lubricity agents or some natural polar materials (refs. 80 and 81). As indicated by figure 9, films of this thickness would represent large increases in surface separation over those of the smaller EHD film

<sup>\*</sup> A possible exception is wear, if corrosion (i.e., chemical reaction involving substrate metal) occurs sufficiently rapidly and the reaction product disperses sufficiently in the bulk lubricant or spalls readily.

<sup>\*\*</sup> Surface separation by monomolecular layers of boundary lubricants is often presumed under effective boundary lubrication conditions; however, careful examination of experimental support suggests that much thicker lubricant layers could (ref. 15) and probably do (ref. 74) exist on the load-supporting asperities. Hence monolayer lubrication is not considered in the present paper.

TABLE 8.—*Heat Transfer Parameters for Surfaces and Lubricant-Type Materials*

Material	Thermal conductivity, $k$ , $10^{-3}$ cal/(cm-sec-°C)	Thermal diffusivity, $\alpha = \frac{k}{c\rho}$ , $10^{-3}$ cm <sup>2</sup> /sec*	Contact coefficient = $\sqrt{k\rho c}$ , $10^{-2}$ cal/(cm <sup>2</sup> -sec <sup>1/2</sup> -°C)*
1% C steel	110	130	31
FeO	1.3	—	—
Fe <sub>2</sub> O <sub>3</sub>	1.4	1.5	4
Liquid and solid hydrocarbons	~0.5	~0.8	~1

\*  $c$  = specific heat,  $\rho$  = density (ref. 76).

TABLE 9.—*Elastic Moduli of Solid-Like Materials*

Material	Young's modulus, d/cm <sup>2</sup>	Ref.
Steel	$2.1 \times 10^{12}$	76
Common inorganic glass	$\sim 7 \times 10^{11}$	77
Polymeric soap glass	$\sim 10^{11}$	78
Organic glass	$\sim 10^{10}$	79
Rubber	$\sim 10^8$	79

thicknesses. In addition to film formation by adsorption of lubricant constituents, it is likely that viscous colloidal films are formed by the interactions of bulk lubricant and reaction products generated at wear surfaces. Iron soaps formed at the wear surfaces are generally insoluble in the bulk lubricant; however, with occluded bulk lubricant a grease-like film might be expected (ref. 68). Solid wear debris in the 100Å size range also should be dispersible (ref. 82). The viscous characteristics of the colloidal structure should depend on concentration and size of the particles and on the presence of surfactive materials such as lubricity agents (ref. 83).

Certainly the colloids would be expected to be non-Newtonian. The flow properties of import to flowing film thickness would be those at the temperatures, pressures, and shear rates of the macro- and micro-conjunction entrances. It appears at this time that the existence of the viscous colloidal layers is supported circumstantially but needs to be demonstrated explicitly (refs. 8, 68, and 84).

The shear stresses produced by flowing lubricant films generally decrease with increased film thickness and with film viscosity. In this case the important viscosity is that at the temperature, pressure, and shear



rate within the film. Any factor tending to reduce film viscosity for a given film thickness should reduce shear stress. Thus paraffinic oils because of their smaller viscosity-pressure coefficients, give lower shear stresses (as measured by friction) than naphthenic oils when sliding velocities are not too large (ref. 85). The low shear stresses observed with lubricity agents at low sliding velocities also may reflect the very low pressure-viscosity coefficients.

At higher sliding velocities, viscous heating is the most important characteristic affecting shear stresses (refs. 69 and 72). For mineral oils (and presumably low thermal-conductivity colloidal films) this becomes important for sliding velocities of the order of a few centimeters per second at the pressures in concentrated conjunctions (ref. 72).<sup>\*</sup> At high sliding velocities, viscous heating dominates and shear stress (i.e., friction) for a constant film thickness and sliding velocity becomes approximately proportional to the ratio of the lubricant thermal conductivity to its temperature coefficient of viscosity. This ratio decreases slightly in going from straight methylene-chain hydrocarbons to naphthenic mineral oils. At high sliding velocities, temperatures in oil films may be a few hundred degrees centigrade higher than those of the confining surfaces (refs. 72 and 86).

Consider now the solid-like films on bearing surfaces and the extent to which they may control flowing lubricant film thickness, asperity stresses, and substrate plastic flow. Table 9 lists thermal properties of materials similar to those involved in the bearing surfaces (ref. 76). Maximum temperature rises on a bearing surface moving relative to a heat-generating conjunction are inversely proportional to the contact coefficient (refs. 77, 86, and 87). The depth to which the heat penetrates the surface is proportional to the square root of the thermal diffusivity (ref. 86).

The properties in table 8 indicate that the order of ten-fold higher temperature rises (i.e., flash temperatures) would be expected on iron oxide and hydrocarbon-type surfaces rather than on steel. Certainly many films are too thin to act thermally as bulk iron oxide or hydrocarbon (refs. 86 and 87). In these cases, surface films increase the flash temperatures somewhat and partially insulate the substrate; however, the theory is extremely complicated. Calculations in Appendix B, based on available theory (ref. 86), indicate that EP films of the order of several  $0.1 \mu\text{m}$

---

<sup>\*</sup> Reference 72 indicates that the heating becomes important when  $\frac{\gamma\eta_p U_s^2}{8k} \approx 1$ ,

in which  $\gamma$  = viscosity-temperature coefficient,  $k$  = film thermal conductivity,  $\eta_p$  = film viscosity at pressure and surface temperature, and  $U_s$  = sliding velocity. For typical mineral oils,  $\frac{\gamma\eta_p U_s^2}{8k}$  is approximately unity when  $\eta_p U_s^2 \sim 10^7 d/\text{sec}$ .



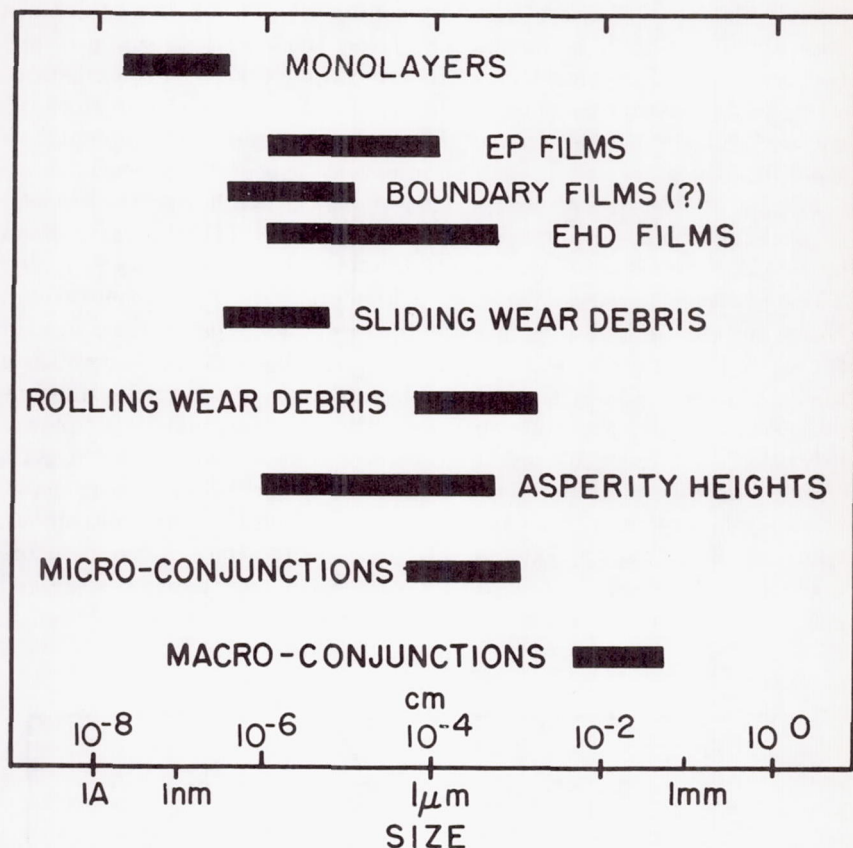


FIGURE 9.—Sizes pertinent to concentrated-conjunction lubrication.

thick should be sufficient at modest sliding velocities to confine almost all the heat to the surface layer between asperities. Thus EP films should effectively insulate the substrate. This insulation in turn should decrease the shear resistance of flowing films and the tendency for thermal transformation and softening of substrate steel. On the other hand, it will put a premium on the thermal shock resistance of the EP film and, by increasing flowing lubricant film temperatures, probably decrease the amount of flowing lubricant entering micro-conjunctions.

In addition to thermally insulating the bearing surfaces, it is important if EP films can also reduce effective elastic moduli at micro-conjunctions. Table 9 lists approximate elastic moduli for materials similar to those involved on bearing surfaces. The high organic content of solid-like EP films plus the usually amorphous nature of their inorganic portions means that they should have elastic moduli lower than steel. Lower elastic

moduli films can reduce stresses at conjunctions if the film thickness approaches the order of magnitude of the Hertzian load-support zone (ref. 88). Thus, for typical (ref. 74) micro-conjunctions between asperities of load-support-area diameter about  $1\text{ }\mu\text{m}$ , EP films of the order of  $0.1\text{ }\mu\text{m}$  and thicker should reduce the Hertzian stress. "Cushioning" of asperities, therefore, is probably an important aspect of EP action.

Fatigue of asperities after repeated stressing would be expected to lead to particles with large dimensions of  $1\text{ to }10\text{ }\mu\text{m}$  (ref. 74); this expectation has been confirmed by wear debris from rolling contact surfaces (ref. 89). These particles are larger than all but the largest EHD film thicknesses. Hence, to escape from a concentrated conjunction, the particles and/or the confining surfaces must be fractured or plastically deformed on a microscale (ref. 90). When sliding is appreciable, the particles must be dragged along the confining surfaces leading to the small-scale "plow," "scratch," or "abrasion" marks observed on the bearing surfaces. As a trapped particle and confining surfaces are stretched by plastic deformation, contaminant films between them must fracture or become thinner. Adhesion would be expected if sufficient clean surface is bared before disintegration of the particle by fracture. Seemingly plowing wear and adhesion could be controlled by: (1) producing particles of sufficiently

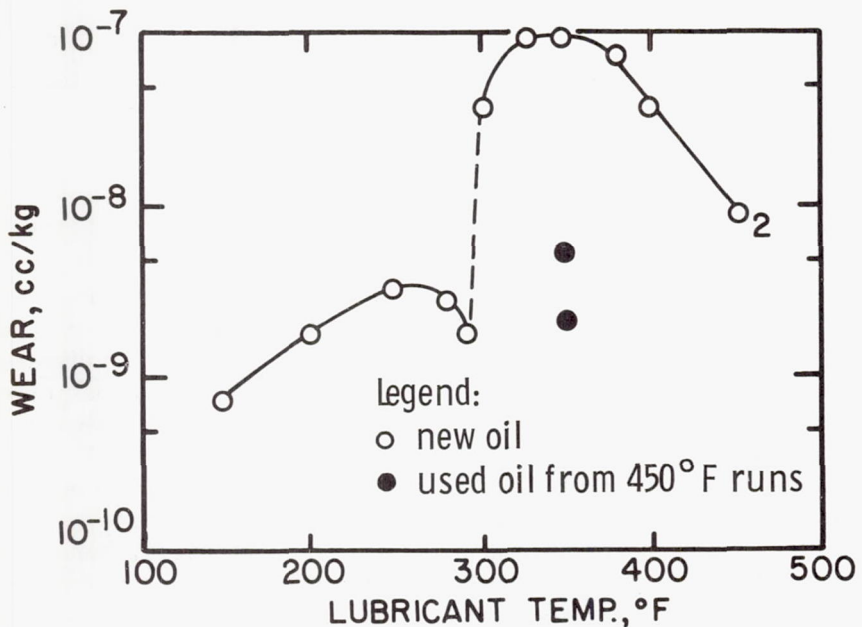


FIGURE 10.—Lubricant run-in. FBM, 90 kg, 9.23 rpm, 11 min. Lubricant: MIL-L-2105B.

low hardness to avoid plastic deformation of the confining surfaces, and/or (2) causing brittle shattering of particles. Low-hardness particles seem to be an accepted characteristic of EP film fragments (ref. 70), while Rehbinder-effect embrittlement of metal particles and surfaces by EP agents and oxygen also seem likely (ref. 74).

Data from many investigators are consistent with the hypothesis that the Rehbinder effect of environment on ductility generally might be expected to help control the extent of adhesive surface disruption above seizure. Lubricity agents, which usually make metals more ductile (ref. 91), increase surface damage; oxygen and EP agents, which usually make metal less ductile, reduce surface damage (ref. 74). The 100-kg wear data in table 4 show these effects of lubricity and EP agents.

#### INTERACTIONS OF CHEMICAL AND PHYSICAL MECHANISMS

The foregoing summary of chemical and physical mechanisms indicates that several of each kind of mechanism will generally be occurring both concurrently and consecutively. These interactions among the various mechanisms can cause both reinforcement and interference. Thus energy dissipation in the flowing lubricant film raises the temperature, leading to chemical reactions giving products which have desirable flow properties on the one hand, or which are too corrosive on the other. As illustrated in figure 10, exposing a lubricant to sliding at high temperature may improve the lubricant, presumably because of a catalytic effect of the sliding surfaces or high stresses since sliding was necessary but frictional heating was negligible.\*

Chemical reaction products formed at wear surfaces can be either beneficial or harmful because of their physical and chemical properties. Hydrocarbons can react at or on wear surfaces to yield a friction polymer which by virtue of its viscous effect provides effective hydrodynamic lubrication. This effective lubrication may be prevented by too much solubilization of polymer by bulk lubricant (ref. 33) or by inhibition of polymer formation. The latter results from too much oxygen with comparatively nonreactive naphthenes and leads to metal-oxide rather than friction-polymer wear products (ref. 33); it also results from too little oxygen with very reactive aromatics which are torn apart by the metal and lead to metal carbide (ref. 8).

In general, the extent of wear and surface damage in a concentrated conjunction depends on a balance among the various chemical and physical steps. The effects of the balance can be quite large, as illustrated by figure 11; in this case, wear\*\* with lubricants of varying constitution

\* Randall, T. H.; and Fein, R. S.: Unpublished, 1961.

\*\* Wear in figure 11 is in terms of wear coefficient. Wear coefficient = (wear volume)/(hardness)/(normal load) (sliding distance).



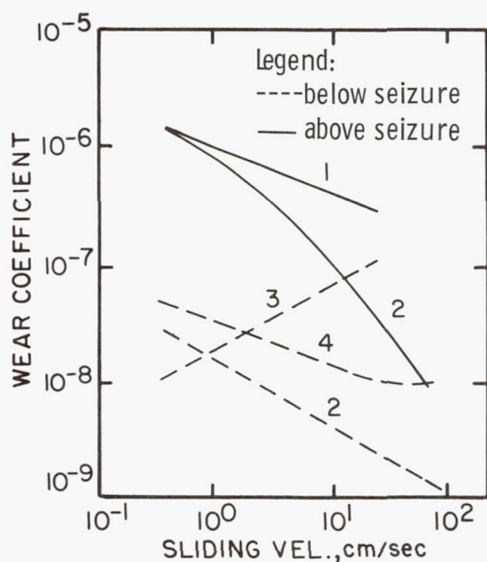


FIGURE 11.—Effect of sliding velocity and constitution on wear. FBM. Lubricant type; 1, squalane; 2, 0.43-percent stearic acid/squalane; 3, 3-percent tribenzylboarate/cetane; 4, MIL-L-2105B.

responds differently to physical conditions imposed by the operating variable of sliding velocity. Other cases of chemical and physical step balance abound (refs. 92 to 94). The balance among chemical and physical factors is often delicate, as evidenced by the abrupt transitions (e.g., seizure or welding) in wear and surface damage that can occur with small changes in operating and constitutional variables. Table 10 shows an example of how both constitutional and physical factors can affect the delicate balance (ref. 95). Clearly understanding the balances of practical import requires much further elucidation of the nature of the chemical and the physical steps and their interactions.

#### NEEDS FOR FURTHER WORK

The preceding sections referred to the profound effects on wear and surface damage of constitutional and operational variables. They also summarized present knowledge concerning the nature of chemical reaction products formed on bearing surfaces and the steps involved in the formation and action of these products in wear and surface-damage control. These summaries pointed out that actual wear and surface-damage control must depend on the physical properties of reaction-product-covered

TABLE 10.—*Effect of Additive and Surface Finish Interaction on Load-Carrying Capacity\**

Additive in mineral oil	Failure load, lb	
	1 to 2 $\mu$ -inch	10 to 15 $\mu$ -inch
None (I)	111	44
7% methyloleate (L)	150	64
7% sulfurized methyloleate (EP)	208	140
S-Cl-P type (EP)	255	395
S-Cl type (EP)	282	417

\* SAE machine, 500 rpm. Lubricant type: I, indifferent; L, lubricity; EP, extreme pressure (ref. 95).

bearing surfaces. It was also indicated that these properties will influence the formation of subsequent reaction products. Actual measurements of the physical properties of bearing surfaces with in-situ reaction products are almost entirely lacking despite their paramount importance. Hence these measurements appear to be the greatest immediate research need.

The accomplishment of these measurements requires appreciable technique development. The reaction product properties that should be measured are the rheological properties of effective elastic modulus and flow resistance (e.g., viscosity) plus the thermal properties. The small thickness of reaction product layers pertinent to the lubrication, wear, and surface damage, and the apparently asperity-scale distribution of reaction products on bearing surfaces, probably require use of small lightly-loaded probes for property measurement. Clues to the feasibility of measurements of this kind are anomalously high EP film micro-hardnesses (ref. 41), and apparent reduction of surface roughness by oils (ref. 96). In the former case, the small micro-indentations probably resulted from the low EP-film elastic modulus; in the latter case, the smoothing probably resulted from more favorable geometry for micro-EHD film formation under the stylus in the valleys between asperities.

When suitable techniques for measuring the pertinent physical characteristics of surfaces are developed, correlations are needed relating these characteristics with constitutional and operational variables and with the accompanying friction, wear, and surface damage. Such correlations should permit deduction of the important physical mechanisms and their relationship to constitution. These correlations and deductions may permit semi-quantitative predictions of the interacting effects of chemical and physical variables. The complexities of interactions probably preclude predictions as quantitatively accurate as those achieved for the purely physical process of EHD lubrication.

## ACKNOWLEDGMENT

Dr. K. L. Kreuz, through years of close stimulating collaboration with the author, is responsible for many of the views presented in this paper.

## APPENDIX A. SURFACE FILM EFFECT

Consider two plane surfaces moving parallel to each other at a velocity  $U_s$ . If each surface is uniformly covered with a liquid film of thickness  $d$ , and the surface films are separated by a liquid layer of thickness  $h$ , the total surface separation is  $h+2d$ . Equating viscous shear stresses across the surface film of viscosity  $\eta_f$  and central or bulk film of viscosity  $\eta$ ,

$$\text{Shear stress} = \frac{\eta_f u}{d} = \frac{\eta(U_s - 2u)}{h}$$

Solving for  $u$ , the velocity difference across a surface film, gives

$$u = \frac{U_s}{2 + \frac{\eta_f h}{\eta d}}$$

This equation indicates that the surface film acts like an extension of the surface when  $(\eta_f h)/(\eta d) \gg 2$ ; in this case  $u \rightarrow 0$  and the shear stress approaches

$$\frac{\eta U_s}{h}$$

## APPENDIX B. HEAT PENETRATION INTO SURFACE

The maximum effect of heat transfer at depth  $z$  beneath surface occurs at a time  $\tau_z = z^2/2\alpha$  after the sudden application of a heat source to the surface (ref. 86). In the case of a surface moving with velocity  $U$  through a concentrated conjunction of width  $2b$ , the frictional heat source will be applied to the surface for time  $\tau_b = 2b/U$ . If  $\tau_b \ll \tau_z$ , then the heat will be essentially confined to a depth less than  $z$  during the time the surface is in the conjunction. Thus  $z^2/2\alpha \gg 2b/U$  is a criterion for confinement of an appreciable temperature rise within a surface layer with thickness  $z$ ;

rearranging and taking the square root, this criterion becomes  $z \gg 2\sqrt{\frac{\alpha b}{U}}$ .

The above criterion can be used to estimate the thickness of an EP layer required to insulate the substrate steel. Table 8 indicates that  $\alpha$  for materials similar to EP layers is about  $10^{-3}$  cm<sup>2</sup>/sec. Substituting into the criterion yields  $z \gg 6 \times 10^{-2} \sqrt{\frac{b}{U}}$ . For a macro-conjunction in which  $b$  is



$10^{-2}$  cm, this reduces to  $z \gg 6 \times 10^{-3} / \sqrt{U}$  cm. Obviously with very high velocities (i.e.,  $10^4$  cm/sec) and thick (i.e.,  $\sim 1 \mu\text{m}$ ) EP layers, the EP layer will completely insulate the substrate; however, in most cases, only partial insulation is provided.

Application of the above criterion to micro-conjunctions between asperities requires substitution of the sliding velocity,  $U_s$ , for the surface velocity. For a micro-conjunction,  $b$  is about  $10^{-4}$  cm. An insulating EP film thickness on the asperity is then  $z \gg 6 \times 10^{-4} / \sqrt{U_s}$  cm. Apparently typical EP film thicknesses (i.e.  $\sim 0.1 \mu\text{m}$ ) should be sufficient to insulate the asperity substrate even at modest sliding velocities (i.e.,  $\sim 100$  cm/sec).

### DISCUSSION

H. E. SLINEY (NASA Lewis Research Center, Cleveland, Ohio)

This paper provides an unusually comprehensive and balanced discussion of the phenomena which occur in lubricated concentrated contacts. Although the title indicates the paper is primarily concerned with chemical phenomena, the author makes a clear attempt to indicate where chemical effects fit into the overall picture in regard to both boundary and elastohydrodynamic lubrication. EHD theory is based upon fluid mechanics analyses which take into account the elastic deformations of the bearing surfaces and the effect of pressure on the viscosity of the lubricant. It is difficult to visualize how chemical effects could be incorporated into these analyses. Thus, EHD theory provides a very important, but nevertheless incomplete picture of the important phenomena which can occur within lubricated contact. It would appear, therefore, that EHD effects and chemical effects must be considered in an uncoupled manner and the resulting conclusions synthesized to provide a more complete understanding of conditions within the contact.

The author describes in detail the influence of clean metal surfaces on catalyzing the polymerization and the oxidation of organic lubricants. The polymers and/or organometallics formed by these reactions can disperse within the parent hydrocarbon and increase its effective viscosity in the vicinity of the reaction site (the lubricated contact). The film thickness under these conditions will be greater than predicted by EHD analyses. When one considers the severity of the conditions within the lubricated contact, the occurrence of chemical reactions is not only plausible but seems almost inevitable.

The author briefly discusses the possibility that high pressures might cause crystal transformations in wear products. The discussion of pressure effects can be extended to include pressure-accelerated chemical reactions. It is a well-known principle of chemical equilibrium that increased pressure will accelerate chemical reactions for which the reaction products occupy a smaller volume than the reactants. Cases have been reported

(ref. 97) where the liquid reaction rates increased several hundred times when the pressure was raised from 1 atmosphere to 150 000 psi, although system volume was reduced by only 25 percent. It is of fundamental importance in lubrication that a net volume decrease occurs when new covalent bonds are found in organic reactions. Because of this effect, many liquid phase polymerizations and copolymerizations are strongly pressure-accelerated; these reactions have been studied at pressures up to 75 000 to 150 000 psi, pressures equivalent to lubricant pressures in many concentrated contacts. One of the more common pressure-accelerated polymerizations is chain growth in olefins, this is of direct consequence to lubrication because the presence of unsaturated compounds is common in many mineral oils. When one also considers the possibilities of frictional activation for polymerization as discussed by Hermance and Egan (ref. 98), by Fein and Kreuz and, by Chaikin (refs. 33 and 58), it is apparent that the lubricated contact can be a veritable "chemical mill."

Because of the polymerization reactions that are known to occur under high static pressures, it is likely that chemical reactions can occur within an EHD film or between the lubricant and the bearing metal even in the complete absence of metal-to-metal contact. If the reaction products disperse throughout the lubricant film, they can easily produce changes in the viscometric properties of the lubricant and produce effects that are unaccountable by EHD theory alone.

An interesting question is whether significant pressure-accelerated reactions can occur during the brief residence time of the lubricant in the contact. Will reaction rates be rapid enough to produce an appreciable effect? High reaction rates will, of course, favor accumulation of reaction products in the vicinity of the lubricated contact. Even at moderate reaction rates, however, it seems possible that reaction products could build up in the stationary boundary layer of lubricant on the bearing surfaces.

It is apparent that considerations of lubricated concentrated contacts must be made with an open mind. One should not be so enamored with a successful, important, and popular fluid mechanics approach (EHD) that he forgets chemical and other effects. On the other hand, the proponents of the importance of chemistry in concentrated-conjunction lubrication obviously cannot afford to ignore the importance of hydrodynamic effects in thin film lubrication.

#### LECTURER'S CLOSURE

Mr. Sliney points out that, in agreement with LeChatelier's principle, increased pressure shifts chemical equilibria towards smaller volume constituents. In other words, pressure increases the ratio of the formation rate of smaller volume products (from larger volume reactants) to the rate of the reverse reaction.



Although the ratio of rates is shifted by pressure, individual rates probably decrease because pressure decreases molecular mobility (e.g., pressure effect on viscosity). Abundant evidence for these mobility effects is the common occurrence of the metastable states of solids and viscous liquids (including glasses). These immobile materials frequently contain thermodynamically unstable (i.e., nonequilibrium) concentrations of chemical compounds, crystal structures, vacancies, electrical charges, etc.

In view of the usually slow reactions of immobile systems and the expected effect of pressure on mobility in concentrated contacts, it appears that "normal" reaction rates are too slow to account for the reactions which occur. This view is supported by the negligible chemical reaction observed in almost pure rolling elastohydrodynamic contacts (provided bulk temperatures are low). Consequently, the paper discusses activation of the chemical reactions by the sliding processes.

Although not mentioned explicitly, the activation mechanisms fall into two general classes. First, there are mechanisms which accelerate reactions by increasing the probability of reaching the top of the normal activation energy barrier (e.g., temperature and shear stress). Second, there are mechanisms which provide alternate reaction paths with lower activation energy barriers (e.g., catalysis and exo-electrons). Most likely, both classes of activation mechanisms usually operate simultaneously.

#### REFERENCES

1. SIMHARD, G. L.; RUSSELL, L. W.; AND NELSON, H. R.: Extreme Pressure Lubricants—Film-Forming Action of Lead Naphthenate and Free Sulfur. *Ind. Eng. Chem.*, vol. 33, 1941, pp. 1352-1359.
2. BEECK, O.; GIVENS, J. W.; AND SMITH, A. E.: On the Mechanics of Boundary Lubrication I. The Action of Long Chain Polar Compound. *Proc. Roy. Soc. (London)*, vol. A177, 1940, pp. 90-102.
3. LARSEN, R. G.; AND PERRY, G. L.: Chemical Aspects of Wear and Friction. *Mechanical Wear*, ASM, ch. 5, 1950.
4. BONDI, A.: *Physical Chemistry of Lubricating Oils*. Rheinhold Publishing Corp., 1951.
5. ROELANDS, C. J. R.: Correlational Aspects of the Viscosity-Temperature-Pressure Relationship of Lubricating Oils. Doctoral Thesis, Delft Technical University, 1966.
6. ROUNDS, F. G.: Effects of Base Oil Viscosity and Type on Bearing Ball Fatigue. *ASLE Trans.*, vol. 5, 1962, pp. 172-182.
7. BISSON, E. E.; AND ANDERSON, W. J.: Advanced Bearing Technology. NASA SP-38, ch. 12, 1964.
8. FEIN, R. S.; AND KREUZ, K. L.: Chemistry of Boundary Lubrication of Steel by Hydrocarbons. *ASLE Trans.*, vol. 8, 1965, pp. 29-38.
9. OWENS, R. S.; AND BARNES, W. J.: The Use of Unsaturated Hydrocarbons as Boundary Lubricants for Stainless Steel. *ASLE Trans.*, vol. 10, 1967, pp. 77-84.
10. VINOGRADOV, G.; KOREPOVA, I. V.; PODOLSKY, Y. Y.; AND PAVLOVSKAYA, N. T.: Effect of Oxidation on Boundary Friction of Steel in Hydrocarbon Media and Critical Friction Duties Under which Cold and Hot Seizure (or Welding) Develop. *Trans. ASME, J. Basic Engrg.*, vol. 87, 1965, pp. 741-745.



11. FEIN, R. S.; AND KREUZ, K. L.: Fundamentals of Wear. *Lubrication*, vol. 42, 1956, pp. 149-160.
12. KLAUS, E. E.; AND BIEBER, H. E.: Effect of Some Physical and Chemical Properties of Lubricants on Boundary Lubrication. *ASLE Trans.*, vol. 7, 1964, pp. 1-10.
13. BOWDEN, F. P.; GREGORY, J. N.; AND TABOR, D.: Lubrication of Metal Surfaces by Fatty Acids. *Nature*, vol. 156, 1945, pp. 97-101.
14. BOWDEN, F. P.; AND TABOR, D.: *The Friction and Lubrication of Solids*, Oxford Univ. Press, 1954.
15. FEIN, R. S.: Effects of Lubricants on Transition Temperature. *ASLE Trans.*, vol. 8, 1965, pp. 59-68.
16. KRUSHCHOV, M. M.; AND MATVEEVSKII, R. M.: *Vestnik Mashinostroeniya*, vol. 34, 1954, pp. 12-18.
17. ROUNDS, F. G.: Some Effects of Additives on Rolling Contact Fatigue. *ASLE Trans.*, vol. 10, 1967, pp. 243-255.
18. BORSOFF, V. N.; AND WAGNER, C. D.: Studies of Formation and Behavior of an Extreme Pressure Film. *Lub. Engrg.*, vol. 13, 1957, pp. 91-99.
19. WIGUIST, R. C.; TWISS, S. B.; AND LOESSER, E. H.: Distribution of an EP Film on Wear Surfaces. *ASLE Trans.*, vol. 3, 1960, pp. 40-47.
20. ROUNDS, F. G.: Some Environmental Factors Affecting Surface Coating Formation with Lubricating Oil Additives. *ASLE Trans.*, vol. 9, 1966, pp. 88-100.
21. BEZBOROD'KO, M. D.; PAVOLSKAYA, G. V.; AND TSURKAN, I. G.: *Izvest. Akad. Nauk. SSSR, Otdel. Tech. Nauk.*, vol. 12, 1958, pp. 104-114.
22. VINOGRADOV, G. V.; PAVLOVSKAYA, N. T.; AND PADOLSKU, Y. Y.: Investigation of the Lubrication Process Under Heavy Friction Conditions. *Wear*, vol. 6, 1963, pp. 202-205.
23. TAO, F. F.: The Role of Diffusion in Corrosive Wear. *ASLE Trans.*, vol. 11, 1968, pp. 211-230.
24. GODFREY, D.: The Effect of High Temperature on Friction and Wear. *Proc. 5th World Pet. Cong.*, New York, paper 23, 1959.
25. ROWE, C. N.; FEIN, R. S.; AND KREUZ, K. L.: Relationship Between Lubricant Oxidation and Wear. *ACS Div. of Pet. Chem. Preprints*, vol. 3, no. 4, 1958, A111-A128.
26. KOVED, I.: The Effect of Three Mineral Base Oils on Roller Bearing Fatigue Life. *ASLE Trans.*, vol. 9, 1966, pp. 222-228.
27. APPELDOORN, J. K.; AND TAO, F. F.: The Lubricity Characteristics of Heavy Aromatics. *ACS Div. of Pet. Chem. Preprints*, vol. 13, no. 2, 1960, pp. B15-B25.
28. TINGLE, E. D.: Influence of Water on the Lubrication of Metals. *Nature*, vol. 160, 1947, p. 710.
29. SCHATZBERG, P.; AND FELSEN, I. M.: The Influence of Water and Oxygen on Rolling Contact Lubrication. *ACS Div. of Pet. Chem. Preprints*, vol. 13, no. 2, 1968, pp. B49-B60.
30. APPELDOORN, J. K.: Discussion of Some Effects of Nonhydrocarbon Base Oil Constituents on the Friction and Surface Coating Formation with Three Additives by F. G. Rounds. *ASLE Trans.*, vol. 11, 1968, p. 29.
31. ROUNDS, F. G.: Some Effects of Nonhydrocarbon Base Oil Constituents on the Friction and Surface Coating Formation with Three Additives. *ASLE Trans.*, vol. 11, 1968, pp. 19-30.
32. FOX, H. W.; HARE, E. F.; AND ZISMAN, W. A.: Wetting Properties of Organic Liquids on High Energy Surfaces. *J. Phys. Chem.*, vol. 59, 1955, pp. 1097-1106.
33. FEIN, R. S.; AND KREUZ, K. L.: Cyclohexane Vapor Lubrication of Steel. *ACS Div. of Pet. Chem. Preprints*, vol. 13, no. 2., 1968, pp. B27-B38.
34. GODFREY, D.: Chemical Changes in Steel Surfaces During Extreme Pressure

- Lubrication. ASLE Trans., vol. 5, 1962, pp. 57-66.
35. MEAD, W. L.; REID, W. K.; AND SILVER, H. B.: The Mass Spectra of Basic Zinc Carboxylates. Chemical Communications, vol. 573, 1968.
  36. LUBARSKII, I. M.; LIUBCHEMKO, A. P.; AND NESTERENKO, V. G.: On the Action of Sulphurized Lubricant. Friction and Wear in Machinery (English translation by ASME), vol. 12, 1960, pp. 285-293.
  37. BARCROFT, F. T.; AND DANIEL, S. G.: The Action of Neutral Organic Phosphates as EP Additives. Trans. ASME, J. Basic Engrg., vol. 87, 1965, pp. 761-770.
  38. BIEBER, H. E.; KLAUS, E. E.; AND TEWKSBURY, E. J.: A Study of Tricresyl Phosphate as an Additive for Boundary Lubrication. ASLE Trans., vol. 11, 1968, pp. 155-161.
  39. GODFREY, D.: The Lubrication Mechanism of Tricresyl Phosphate on Steel. ASLE Trans., vol. 8, 1965, pp. 1-11.
  40. FENG, I. M.; PERILSTEIN, W. L.; AND ADAMS, M. R.: Solid Film Deposition and Non-Sacrificial Boundary Lubrication. ASLE Trans., vol. 6, 1963, pp. 60-66.
  41. KREUZ, K. L.; FEIN, R. S.; AND DUNDY, M.: EP Film from Borate Lubricants. ASLE Trans., vol. 10, 1967, pp. 67-76.
  42. CAMPBELL, R. B.: Study of Hypoid-Gear Lubrication Using Radioactive Tracers. Lubrication and Wear Convention, Inst. Mech. Engrs., paper 25, May 23-25, 1963.
  43. BARCROFT, F. T.; AND DANIEL, S. G.: Extreme-Pressure Film Formation on Hypoid Gears—Studies with Sulfur-35. Lubrication and Wear Third Convention, Inst. Mech. Engrs., paper 17, May 27-29, 1965.
  44. MANTEUFFEL, A. A.; YATES, K. P.; BICKFORD, H. J.; AND WOLFRAM, G.: Radiotracers Reveal Activity of Extreme-Pressure Additives in Lubrication. ASLE Trans., vol. 7, 1964, pp. 249-256.
  45. ROUNDS, F. G.: Discussion of The Load-Carrying Mechanism of Organic Sulfur Compounds—Application of Electron Probe Microanalysis by K. G. Allum and E. S. Forbes. ASLE Trans., vol. 11, 1968, pp. 173-174.
  46. FEIN, R. S.; AND RANDALL, T. H.: Discussion of Chemical Changes in Steel Surfaces During Extreme Pressure Lubrication by D. Godfrey. ASLE Trans., vol. 5, 1962, p. 65.
  47. ROUNDS, F. G.: Effects of Additives on the Friction of Steel on Steel I., Surface Topography and Film Composition Studies. ASLE Trans., vol. 7, 1964, pp. 11-23.
  48. FUREY, M. J.; AND KUNC, J. F.: A Radiotracer Approach to the Study of Engine Valve Train Lubrication. Lub. Engrg., vol. 14, 1958, pp. 302-309.
  49. KELLER, D. V.: On the Analysis of Adhesion Data. ASTM-ASLE Conf. on Adhesion and Cold Welding Under Space Conditions, Toronto, May 2, 1967.
  50. EISCHENS, R. P.: Catalysis Studies Related to Boundary Lubrication. Boundary Lubrication—An Appraisal of World Literature, ASME, 1969.
  51. ROWE, C. N.; AND DICKERT, J. J., JR.: The Relation of Antiwear Functions to Thermal Stability and Structure for Metal O, O-Dialkylphosphorodithioates. ASLE Trans., vol. 10, 1967, pp. 85-90.
  52. OWENS, R. S.; ST. PIERRE, L. E.; AND ROBERTS, R. W.: Discussion of Chemistry of Boundary Lubrication of Steel by Hydrocarbons by R. S. Fein and K. L. Kreuz. ASLE Trans., vol. 8, 1965, p. 37.
  53. CAMPBELL, D.; AND A. PETERLIN: Free-Radical Formation in Uniaxially Stressed Nylon. J. Polymer Sci., vol. 6, pt. B, 1968, pp. 481-491.
  54. HAMILTON, G. M.; AND ROBERTSON, W. G.: Lubrication with Oils Containing Polymers. Lubrication and Wear Third Convention, Inst. Mech. Engrs., paper 3, May 27-29, 1965.
  55. BARLOW, A. J.; AND LAMB, J.: The Visco-Elastic Behavior of Lubricating Oils



- Under Cyclic Shearing Stress. *Proc. Roy. Soc. (London)*, vol. A253, 1959, pp. 52-69.
56. COTTRELL, A. H.: Strong Solids. *Proc. Roy. Soc. (London)*, vol. A282, 1964, pp. 2-9.
  57. FRANCIS, S. A.; AND ELLISON, A. H.: Reflection Infrared Studies of Zinc Dialkylphosphorodithioates Films Adsorbed on Metal Surfaces. *J. Chem. and Eng. Data*, vol. 6, 1961, pp. 83-86.
  58. CHAIKIN, S. W.: On Frictional Polymer. *Wear*, vol. 10, 1967, pp. 49-60.
  59. TAMAI, Y.; SUZUKI, M.; AND MOMOSE, Y.: The Effect of Exoelectrons on the Adsorption and the Lubricity of Hydrocarbon Liquids on Steel. *ACS Div. of Pet. Chem. Preprints*, vol. 13, no. 2, 1968, pp. B159-B162.
  60. KLAUS, E. E. et al.: Fluids, Lubricants, Fuels and Related Materials. *Air Force Materials Lab. Tech. Report AFML-TR-67-107*, pt. II, April 1968.
  61. ACHESON, R. J.; AND GALWEY, A. K.: The Thermal Decomposition of Salts of Mellitic Acid Part III. Chromic, Manganous, Cupric and Zinc Mellitates. *J. Chem. Soc.*, 1967, pp. 1167-1173.
  62. SANIN, P. I.; SHEPELEVA, E. S.; MANNIK, A. O.; KLEIMENOV, B. V.: Chemical Modification of Friction Surfaces. *Trans. ASME, J. Basic Engrg.*, vol. 87, 1965, pp. 771-777.
  63. COLOMBO U.; FAGHERAZZI, G.; GUSSARINI, F.; LAZAVECCHIA, G.; SIRONI, G.: Mechanism of Low Temperature Oxidation of Magnetites. *Nature*, vol. 219, 1968, pp. 1036-1037.
  64. RODYMANS, C. J. M.: Reactivity in High-Pressure Transformations. *Reactivity of Solids*, G. M. Schwab, ed., Elsevier Publ., 1965.
  65. TAMAI, Y.: The Interactions of Wear Debris with Fatty Additives in Lubrication. *Wear*, vol. 2, 1959, pp. 228-236.
  66. SAKURAI, T.; AND BABO, T.: Surface Chemical Behavior of Polar Compounds, Nonaqueous Liquids. Dispersing Effect of The Soap Solution. *Wear*, vol. 3, 1960, pp. 286-296.
  67. ST. PIERRE, L. E.; OWENS, R. S.; AND KLINT, R. V.: Chemical Effects in the Boundary Lubrication of Aluminum. *Wear*, vol. 9, 1966, pp. 160-168.
  68. FEIN, R. S.; AND KREUZ, K. L.: Lubrication and Wear. *Lubrication*, vol. 51, 1965, pp. 61-80.
  69. DOWSON, D.; AND HIGGINSON, G. R.: *Elasto-Hydrodynamic Lubrication*, Pergamon Press, 1966.
  70. GODFREY, D.: Boundary Lubrication. *Interdisciplinary Approach to Friction and Wear*, P. M. Ku, ed., NASA SP-181, 1968, pp. 335-384.
  71. TABOR, D.: Function Growth in Metallic Friction: The Role of Combined Stresses and Surface Contamination. *Proc. Roy. Soc. (London)*, vol. A251, 1959, pp. 278-293.
  72. CROOK, A. W.: The Lubrication of Rollers III. A Theoretical Discussion of Friction and the Temperatures in the Oil Film. *Proc. Phil. Trans. Roy. Soc. (London)*, vol. A254, 1961, pp. 237-258.
  73. GRUNTFEST, I. J.: Apparent Departures from Newtonian Behavior in Liquids Caused by Viscous Heating. *Trans. Soc. Rheol.*, vol. 9, 1965, pp. 425-441.
  74. FEIN, R. S.; AND KREUZ, K. L.: Discussion on Boundary Lubrication. *Interdisciplinary Approach to Friction and Wear*, P. M. Ku, ed., NASA SP-181, 1968, pp. 358-376.
  75. KRAGELSKII, I. V.: *Friction and Wear*. Butterworths, 1965.
  76. LANGE, N. A.: *Handbook of Chemistry*, Handbook Publishers, 1941.
  77. ZWIKKER, C.: *Physical Properties of Solid Materials*. Interscience, 1954.
  78. FITZGERALD, W. E.; AND NIELSEN, L. E.: Viscoelastic Properties of the Salts of Some Polymeric Acids. *Proc. Roy. Soc. (London)*, vol. A282, 1964, pp. 137-146.



79. STAVERMAN, A. J.: Mechanical Properties of Polymers. *Proc. Roy. Soc. (London)*, vol. A282, 1964, pp. 115-120.
80. DERJAGUIN, B. V.; AND KARASSEN, V. V.: Viscosity Studies of Boundary Layers by the Blow-Off Method. *Proc. 2nd International Cong. of Surface Activity*, vol. III, 1957, pp. 531-538.
81. DERYAGIN, B. V.; KARASEV, V. V.; ZAKHAVAIEVA, N. N.; AND LAZAREV, V. P.: Mechanism of Boundary Lubrication and Properties of Lubricating Film—Short- and Long-Range Action in Theory of Boundary Lubrication. *Wear*, vol. 1, 1958, pp. 277-290.
82. ADAMSON, A. W.: *Physical Chemistry of Surfaces*. Interscience, New York, 1960.
83. GOLDSMITH, H. L.; AND MASON, S. G.: *The Microrheology of Dispersions*. Rheology, Academic Press, ch. 2, vol. 4, 1967.
84. BRIX, V. H.: An Electrical Study of Boundary Lubrication—Further Rolls-Royce Investigation Tending to Confirm Deductions from Previous Test. *Aircraft Engrg.*, vol. 19, 1947, pp. 294-297.
85. ROUNDS, F. G.: Influence of Lubricant Composition on the Friction of Steel on Steel. *ACS Div. of Pet. Chem. Preprints*, vol. 3, no. 4, 1958, pp. A129-A151.
86. ARCHARD, J. F.: The Temperature of Rubbing Surfaces. *Wear*, vol. 2, 1958-59, pp. 438-455.
87. JAEGER, J. C.: Moving Sources of Heat and the Temperature at Sliding Contacts. *Proc. Roy. Soc. NSW*, vol. 6, 1943, pp. 203-224.
88. WATERS, N. W.: The Indentation of Thin Rubber Sheets by Spherical Indenters. *Brit. J. Appl. Phys.*, vol. 16, 1965, pp. 557-562.
89. TALLIAN, T. E.; BRADY, E. F.; MCCOOL, J. L.; AND SIBLEY, L. B.: Lubricant Film Thickness and Wear in Rolling Point Contact. *ASLE Trans.*, vol. 8, 1965, pp. 411-424.
90. GIOLMAS, S. N.; AND HALLING, J.: Role of Wear Debris in the Wear Characteristics of a Rolling Element Subjected to Tangential Surface Traction. *Lubrication and Wear Third Convention, Inst. Mech. Engrs.*, paper 16, May 27-29, 1965.
91. KRAMER, I. R.; AND DEMER, L. J.: Effects of Environment on Mechanical Properties of Metal. *Progress in Materials Sci.*, vol. 9, 1961, pp. 133-199.
92. DAVIES, C. B.: Influence of Roughness and Oxidation on Wear of Lubricated Sliding Metal Surfaces. *Annals N.Y. Acad. Sci.*, vol. 53., 1951, pp. 919-935.
93. BENNETT, P. A.: A Surface Effect Associated with the Use of Oils Containing Zinc Dialkyl Dithiophosphate. *ASLE Trans.*, vol. 2, 1959, pp. 78-90.
94. ALLUM, K. G., AND FORBES, E. S.: The Load-Carrying Mechanism of Organic Sulfur Compounds—Application of Electron Probe Microanalysis. *ASLE Trans.*, vol. 11, 1968, pp. 162-175.
95. ASSEFF, P. A.: Discussion of Study of Corrosivity and Correlation Between Chemical Reactivity and Load-Carrying Capacity of Oils Containing Extreme Pressure Agents by T. Sakurai and K. Sato. *ASLE Trans.*, vol. 9, 1966, p. 86.
96. DERYAGIN, B. V.; AND PICHUGIN, E. F.: Boundary Viscosity and Boundary Process in Lubricant Films. *Proc. 2nd An. Union Conf. Friction and Wear in Machines*, *Izd. Akad. Nauk SSSR*, vol. 2, 1959, p. 271.
97. WEALE, K. E.: *Chemistry under Pressure*. Industrial Research, June 1969, p. 50.
98. HERMANCE, H. W.; AND EGAN, T. T.: Organic Deposits on Precious Metal Contacts. *Bell System Tech. J.*, vol. 37, 1958, p. 739.

**Page intentionally left blank**

# Bearing Design Considerations

**F. W. WELLONS**

**T. A. HARRIS**

**SKF Industries, Inc.**

**King of Prussia, Pennsylvania**

This paper discusses the importance of elastohydrodynamic lubrication on rolling-element bearing performance and design. Special emphasis is given to the problems of fatigue and skidding in rolling-element bearings, showing how the knowledge of elastohydrodynamics can be applied to facilitate understanding of these problems and to provide rational design guidance.

## LOAD RATINGS

Together with bearing bore, width, and outside diameter (standardized boundary dimensions), a rolling bearing is usually selected according to a load rating obtainable from the manufacturer's catalog. Most ball and roller bearing manufacturers utilize the Anti-Friction Bearing Manufacturers' Association (AFBMA) Standards (refs. 1 and 2) to establish these catalog load ratings. Contemporary load rating concepts are based on the work of Lundberg and Palmgren (refs. 3 and 4) who coined the term "basic dynamic capacity."

Because a statically applied load on a rotating bearing does not seem to engender a dynamic condition, the AFBMA adopted the simpler, if less descriptive term, basic load rating. Basic load rating refers to that load which 90 percent of a group of apparently identical rolling bearings subjected to similar conditions of load, speed, lubrication, etc., will endure with a fatigue life of 1 million revolutions of the bearing inner ring.

Lundberg and Palmgren related the basic load rating to a fatigue failure consisting of the first occurrence of a spall in a load-carrying surface. They hypothesized that failure commenced as a minute crack emanating from a material weak point slightly below the surface of a rolling member. The following expressions were developed for basic load rating of radial ball bearings and radial roller bearings:

$$C = f_c (i \cos \alpha)^{0.7} Z^{2/3} D^{1.8} \quad (1)$$



$$C = f_c (i \ell \cos \alpha)^{7/9} Z^{3/4} D^{29/27} \quad (2)$$

Values of  $f_c$ , which are a function of bearing geometry and a basic steel, are obtained from tabular data in references 1 and 2.

To determine the rating life of a rolling bearing in a given application, Lundberg and Palmgren further developed the following load-life equation:

$$L_{10} = (C/P)^n \quad (3)$$

where  $n=3$  for ball bearings and  $10/3$  for roller bearings.

The rating life as determined by equation (3) is the fatigue life that 90 percent of the bearings in a given application will endure. Most bearing users do not realize that a quotation of bearing fatigue life is a statistical phenomenon. Moreover, the 90-percent reliability associated with this life is generally much less than that desired by the user.

In equation (3), the equivalent radial load  $P$  is defined as follows:

$$P = XVF_r + YF_a \quad (4)$$

Values of  $X$ ,  $V$ , and  $Y$  based on bearing geometry can be obtained from formulas in references 1 and 2.

#### LIMITATIONS OF THE STANDARDS

The Standards are simplifications of the Lundberg and Palmgren equations. They may be applied to the majority of rolling bearing applications wherein load is not excessive, speed is moderate, and the bearing is mounted normally. The Standards were never intended to be applicable to bearings subjected to conditions of high speed, high temperature, or marginal lubrication. Because most users do not recognize these and other limitations, the Standards are frequently misapplied.

#### LUBRICATION AND LIFE

Those who have earned their gray hair in the bearing industry can recall certain embarrassing installations where chronic and sometimes catastrophic bearing fatigue failures occurred at lives only a mere fraction of that predicted by the established load rating standards. These failures did not appear connected in any pattern, so they were generally written off as freak occurrences and lumped together as lubrication failures. By trial and error, and at considerable expense, it was eventually found that highly loaded bearings turning at very low speeds demanded a much heavier lubricant viscosity than was known to be satisfactory and normal for similar bearings running at higher speeds. Furthermore, although ball and roller bearings were known to carry loads through rolling metal contacts, low viscosity watery-like fluids almost always caused very short-life fatigue failures.

In addition to these lubricant-associated, short-lived situations, it was

also found that bearings with rough ground or otherwise poorly finished raceway or rolling-element surfaces were prone to early failure, where conventionally finished products running with the same lubricant and in the same application would meet or exceed catalog predicted life. All of the major bearing manufacturers were aware of these abnormalities and tacitly accepted these deficiencies in rolling-bearing technology. Fortunately there was generally sufficient know-how present to protect the bearing user from the unknown or to at least fix his application after the first few early failures.

Extensive analytical and laboratory studies of the almost unrecognized lubricant film between the rolling contacts and the deformations of loaded surfaces demonstrated that failures can occur from surface damage prior to any indication of a subsurface fatigue crack. For the practicing engineer these observations of lubricated rolling contacts are best explained by the simple rule that premature surface failures do not occur when the lubricant film between the rolling surfaces is of sufficient thickness to prevent all metal-to-metal contacts. As figure 1 shows, there exist two sets of mountain peaks rolling on each other that should be separated by a lubricant film of predictable thickness.

#### LUBRICANT FILM THICKNESS

It was determined analytically by Grubin (ref. 5) and confirmed experimentally by Sibley and Orcutt (ref. 6) that the lubricant film thickness between rolling surfaces is primarily a function of lubricant viscosity, pressure coefficient of viscosity, bearing geometry, and rolling speed. Harris (ref. 7) simplified the formulation for ball and roller bearings to:

$$\Lambda = H(\mu_o \alpha N)^{0.7} P_o^{-0.09} \quad (5)$$

In equation (5),  $H$  is a function of bearing geometry that can be prepared in graphical format (ref. 7). Lambda ( $\Lambda$ ), a measure of the adequacy of lubrication, is the ratio of minimum lubricant film thickness to the composite rms finish of the "contacting" surfaces.

The significance of the lambda ratio can be understood by examining figure 2. Practically this chart indicates the degree of metallic contact between the rolling element and raceway for a given set of operating conditions recognizing that these highly finished rolling surfaces actually have some degree of measurable roughness, characteristic of a particular bearing configuration and manufacture. The vertical axis shows the percentage of time for which there is no sporadic contact between the extreme peaks of these rolling surfaces. Thus a 100 percent film would indicate no contact at all, where a zero film would indicate a complete lack of lubricant separation. Research and data collection have shown the percent film separation that can be expected for any given lambda ratio. Typical but

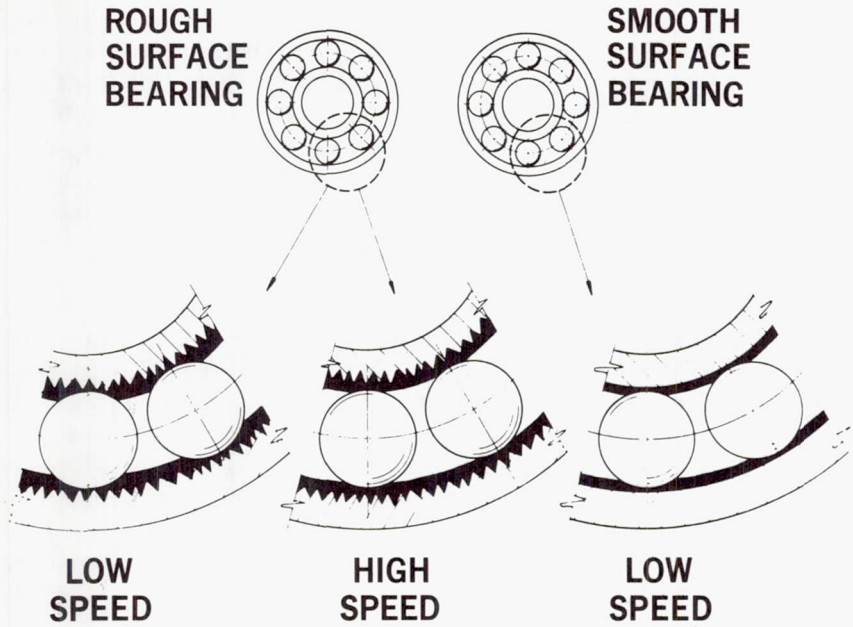


FIGURE 1.—Illustration of the effect of surface roughness on the lubricant film thickness required to prevent metal-to-metal contact.

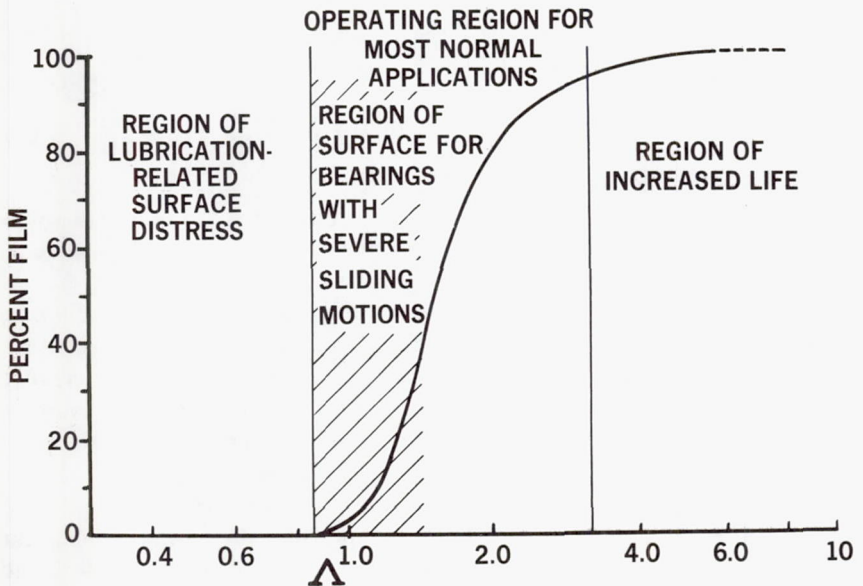


FIGURE 2.—Lubricant film effect on bearing performance.



significant operating regimes have been identified where a rolling bearing might be operating: (1) without any appreciable film separation, (2) where there is measurable but not necessarily complete film separation, characteristic of most applications and in which case normal life is expected; or (3) where the film thickness is four or more times the composite surface roughness where it is known that bearing life can be appreciably longer than predicted by accepted load rating theory.

#### SURFACE-ORIENTED DISTRESS

In addition to the classical bearing surface fatigue failure emanating from minute subsurface imperfections, a new surface phenomenon called a "furrow" has crept into bearing vocabulary. This is a very short, hardly noticeable scratch of insignificant depth in the direction of rolling and somewhere along the periphery of the load-carrying portion of the raceway. Under marginal lubrication, even with the very best finish, we have sporadic film breakdown at these fissures. Such defects under the high contact pressure appear to leak oil, causing metallic contact at the edges. Oddly enough, minor dirt denting or light corrosion that might affect the whole rolling surface does not seem to have the same life-killing effect with comparable weak film support. This is probably because the mechanically induced furrow has very sharp and steep edges at the rolling surface where rolled-in debris and rolled-over corrosion pits are shallow with generally rounded edges. However, severe pitting or mechanical denting have been shown to produce surface-initiated spalls. In the same manner, skidding or smearing within the roller contacts caused by too high accelerations, frozen lubricant, or mounting damage can shorten bearing life due to intermittent metal contacts during subsequent rolling over these jagged surfaces.

#### FILM THICKNESS AND BEARING DESIGN

It is clear from the foregoing discussion that the load rating standards must be modified to include the effect of lubricant film thickness on endurance. Although a qualitative estimator has been established, a definitive quantitative method is yet required. Tallian et al. (ref. 8) developed figure 3 based on the experimental effort conducted to date. The total effort to define EHD phenomena, while it has been extensive, has only scratched the surface with regard to establishing the effect of lubricant film thickness on bearing fatigue life. A simplistic method to include this effect in the load rating procedure is

$$L = a_2 \left[ \frac{C}{P} \right]^n \quad (6)$$

where  $a_2$  is a lubrication-life multiplier taken from data similar to figure

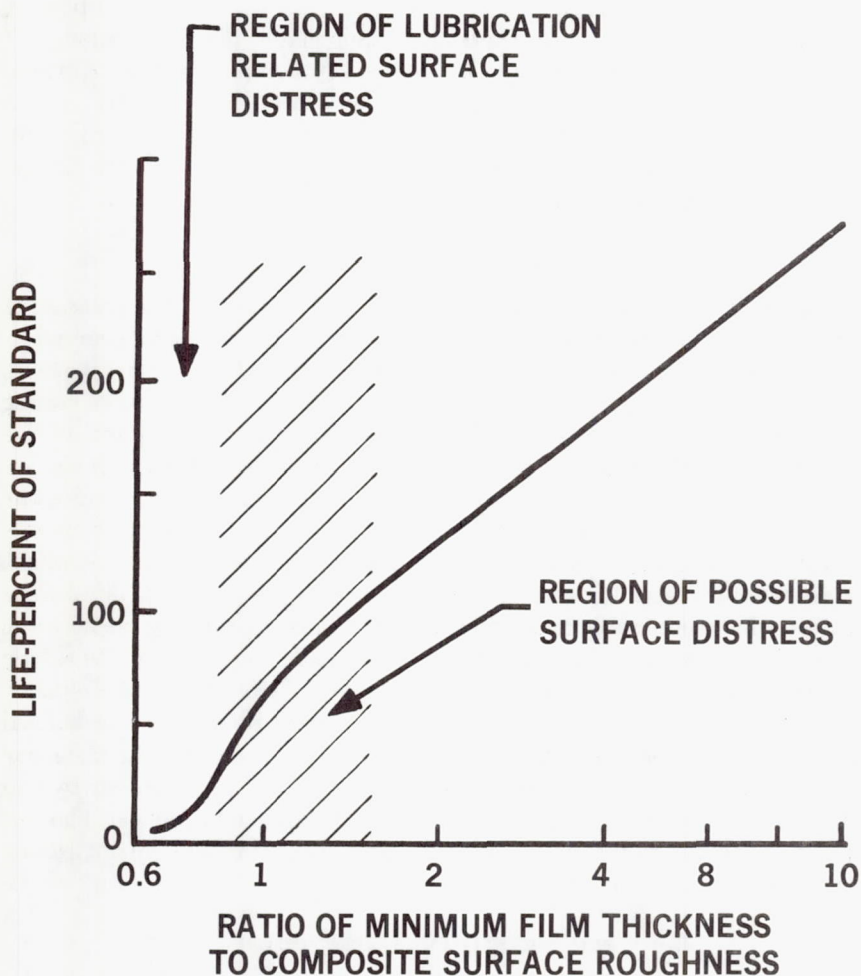


FIGURE 3.—Fatigue life as a function of the minimum lubricant film thickness/roughness ratio.

3. It may be noted that other life multipliers,  $a_1$  and  $a_3$ , may be established for reliability of performance and life improvement owing to bearing material and/or fabrication, respectively.

From equation (5) it can be observed that in a given application superior bearing performance can be achieved by:

- (1) Improving the lubricant viscosity characteristics
- (2) Improving the bearing surface finishes

The former might be achieved by providing cooler lubricant to the bearing

or by changing the lubricant. In the latter case and in a figurative sense, the surfaces involved might be as large as a living room rug while the measuring trace on which the lambda ratio is based would be about as long as three or four big flea hops. Thus integrity of the surface is pretty much dependent upon equipment and processes and not on an inspection method.

Assuming that endurance can be improved by application of either of the foregoing methods, the designer can opt for reduction in bearing size or achievement of greater reliability. Figure 4 (ref. 9) indicates the fatigue life reductions with increased reliability of performance. However, with a life multiplier of two for example, owing to superior lubrication or surface finish, the same endurance can be achieved with a reliability of 96.5 percent in lieu of the standard 90 percent. The latter approach is preferable to reduction in bearing size because of possible misalignment and fitting problems owing to increased flexibility of shaft and housings. It is further evident that, when life improvement due to lubrication is combined with that justified by improvement in bearing steel, the total effect will ultimately lead to a design life that a bearing can attain with 100-percent reliability.

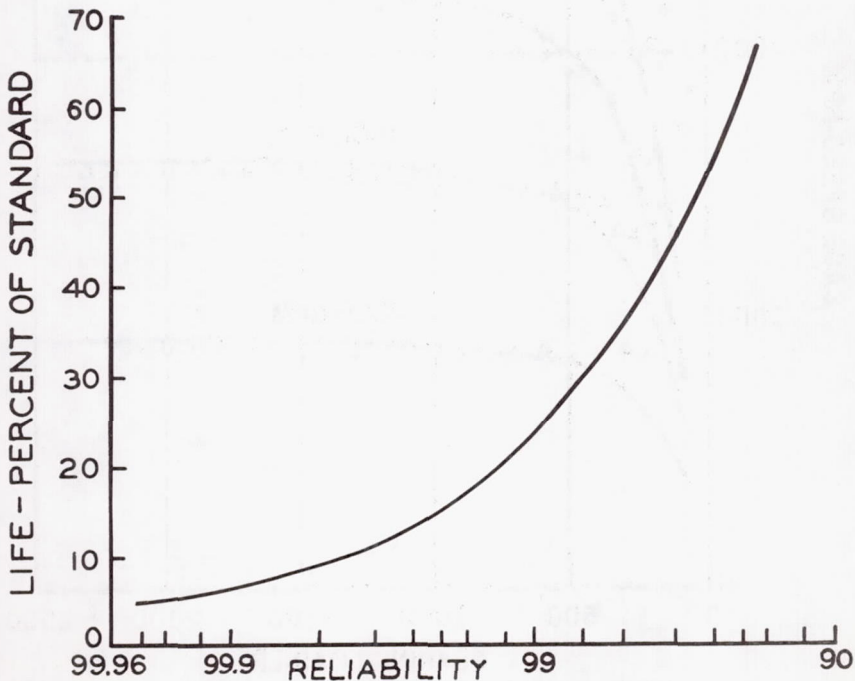


FIGURE 4.—Fatigue life as a function of reliability.



EFFECT OF EHD FRICTION FORCES

Whether a ball or roller bearing can operate with theoretical epicyclic internal speeds is determined by friction forces occurring in the rolling element-to-raceway contacts. In estimating cage and roller speeds in a high-speed cylindrical roller bearing, Harris (ref. 10) was able to estimate the occurrence of gross sliding between the rollers and inner raceways. Figure 5 shows the reduction in cage speed indicative of the "skidding" bearing.

SKIDDING AND ROLLER BEARING DESIGN

To prevent skidding in cylindrical roller bearings, however, it has been found necessary to radially preload bearings by grinding the bearing's

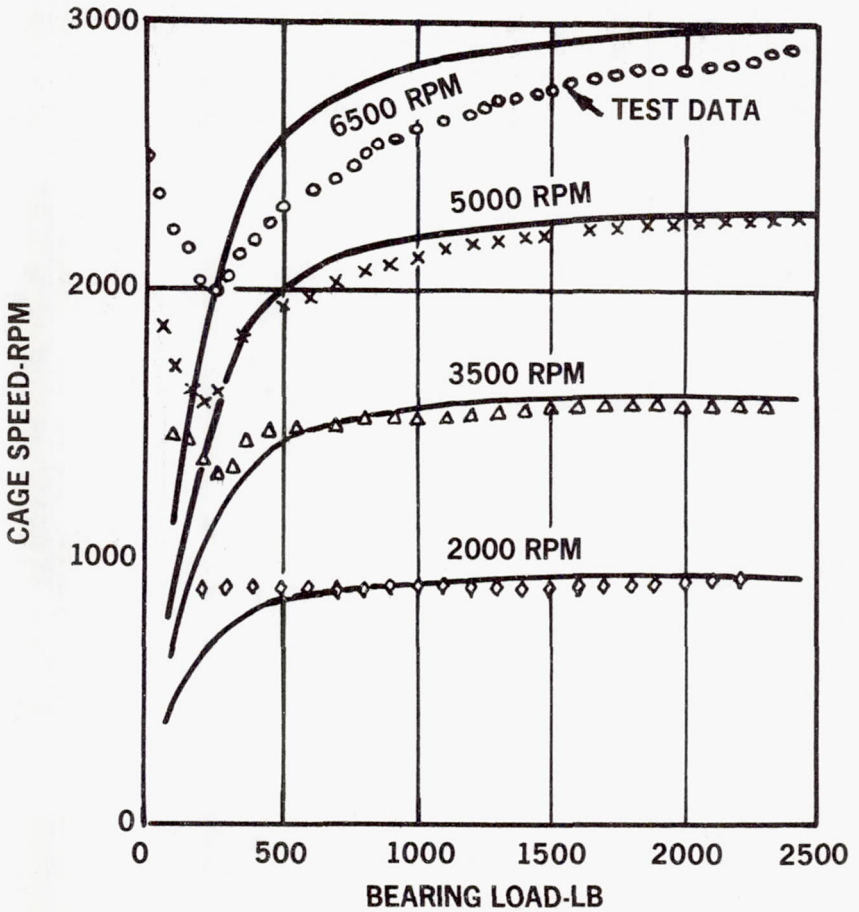


FIGURE 5.—Cage speed vs load and inner-ring speed for a cylindrical roller bearing.

outer raceway elliptical such that its minor axis is smaller than the diameter of the bearing over the rollers by a few thousandths or even hundredths of an inch. This "out-of-round" ring is then squeezed over the roller set, selectively preloading the bearing as shown by figure 6. Determination of the amount of out-of-roundness required to prevent or minimize skidding is a complex process requiring the use of a digital computer program. References 10 to 12 describe in detail the mathematical treatment of the skidding phenomenon in different types of preloaded, cylindrical roller bearings. Figure 7 illustrates the variation of cage speed versus out-of-roundness in a typical application.

Selection of the proper amount of out-of-roundness for a high-speed cylindrical roller bearing is a complex problem. Too much preloading will cause a drastic reduction in fatigue life; too little may cause insignificant reduction in skidding. The problem is further complicated in extremely high-speed applications where expansion of hollow shafts is greatly affected by speed. Hence skidding may be precluded at high speed, but it may occur at somewhat slower speed. Conversely, skidding might be negligible at all speeds; however, fatigue life is greatly reduced at the higher speeds owing to added interference caused by shaft expansion.

With regard to damage of rolling surfaces, the amount of skidding that may be tolerated in a given application appears to be a function of surface finish and heat removal. If the lubricant film thickness generated by the rolling-sliding members is sufficiently greater than the rms composite roughness of the rolling-contact members, then no damage due to skidding may occur. The reverse situation apparently produces surface damage and incipient fatigue. A judicious bearing run-in procedure may further improve the rolling surfaces and thereby ameliorate this situation.

#### ESTIMATION OF FRICTION FORCES

In development of the roller-bearing skidding model, Newtonian lubricant behavior in the roller-to-raceway contacts was assumed; i.e., the lubricant frictional or shear stress obeys the law

$$\tau = \mu \frac{\partial u}{\partial z} \quad (7)$$

Johnson and Cameron (ref. 13) among others demonstrated that only a limited Newtonian regime exists and that lubricant behavior in concentrated load contacts is markedly non-Newtonian as sliding speed assumes appreciable magnitude. Moreover, the thermal effects on viscosity attendant with EHD lubrication (ref. 14) in a contact do not appear to have a significant effect on the non-Newtonian behavior insofar as the friction coefficient is concerned. Perhaps the non-Newtonian lubricant behavior is the reason that the skidding mathematical model is valid

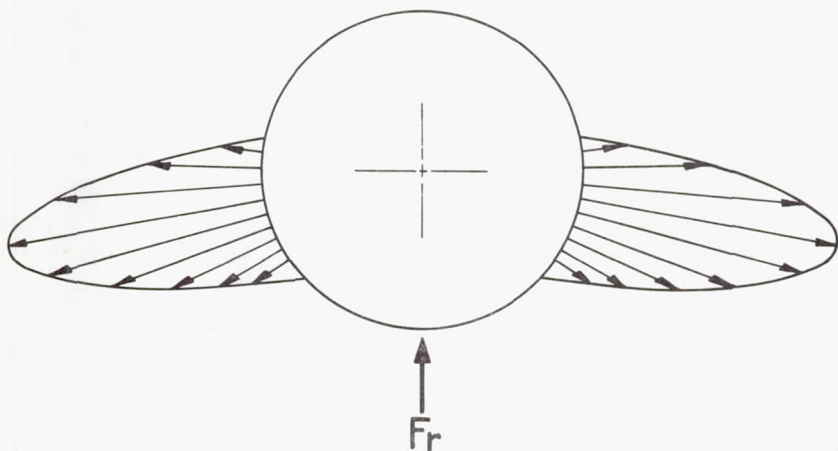


FIGURE 6.—Load distribution in a radially loaded cylindrical roller bearing having an out-of-round outer raceway.

only in the region of relatively small sliding speeds. On the other hand, the current model successfully predicts the onset of skidding and is effectively used to develop anti-skid bearing designs.

#### CAGE-TO-ROLLING ELEMENT LOADING

An interesting spin-off from the inclusion of EHD friction forces in the determination of skidding has been the ability to estimate cage-to-rolling element loading. Figure 8 (taken from reference 10) schematically indicates the forces acting on a roller in the direction of rolling. It is evident that friction forces play a dominant role in the magnitude of the loading applied by the roller to the cage or vice versa. Preliminary analyses assuming Newtonian fluid behavior indicate that such cage web loading is relatively light when:

- (1) Cage-to-riding land friction is negligible
- (2) Viscous drag of the orbiting roller is slight owing to mist-type lubrication
- (3) Misalignment is absent

Figure 9 is a curve typical of such cage-to-roller loading. It is further apparent that if cage-to-roller loading can be predicted for any given application, it is possible to design the bearing to minimize such loading or a bearing cage to withstand it. Since cage destruction is a prime cause of high-speed bearing failure, research directed toward an accurate and universal definition of shear stress (or coefficient of friction) in concentrated load contacts would be well justified.



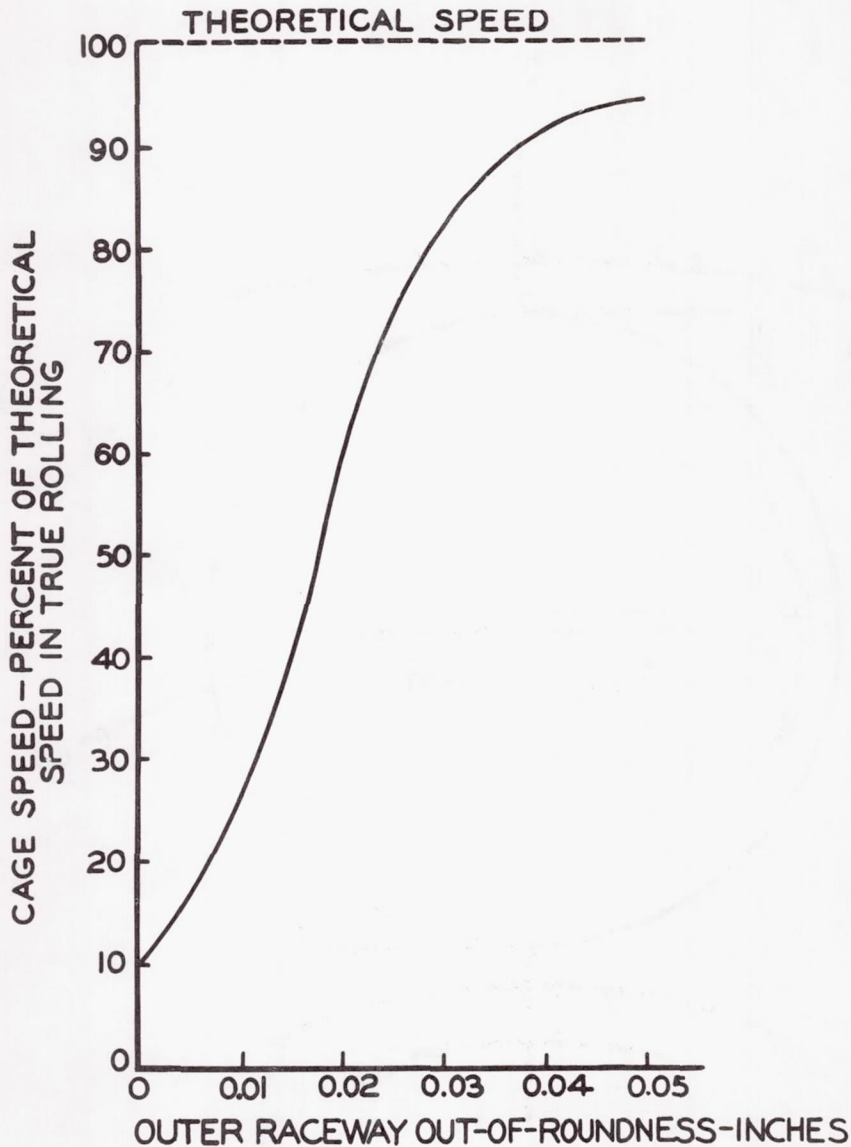


FIGURE 7.—Cage speed vs out-of-roundness for a high-speed cylindrical roller bearing subjected to negligible radial loading.

#### BALL BEARINGS AND RACEWAY CONTROL

To determine the ball speeds of rotation in high-speed ball bearings, it has been considered necessary to assume a condition of raceway control with respect to ball motion. A ball was assumed to roll and spin simul-

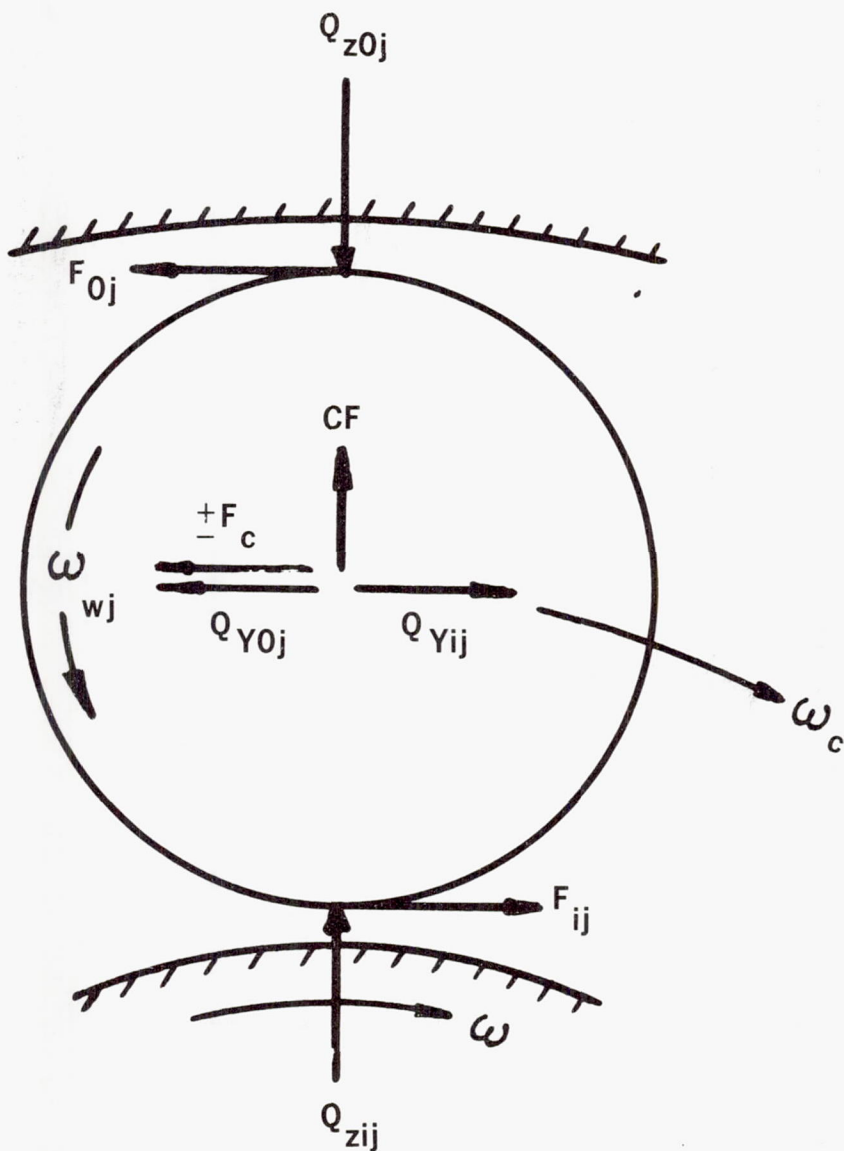


FIGURE 8.—Forces acting on a roller.

taneously relative to one raceway and merely roll relative to the other. Furthermore, only the frictional contact at the controlling raceway was presumed to resist the gyroscopic moment, and ball rotation owing to the gyroscopic moment was not assumed to exist. Using the EHD friction

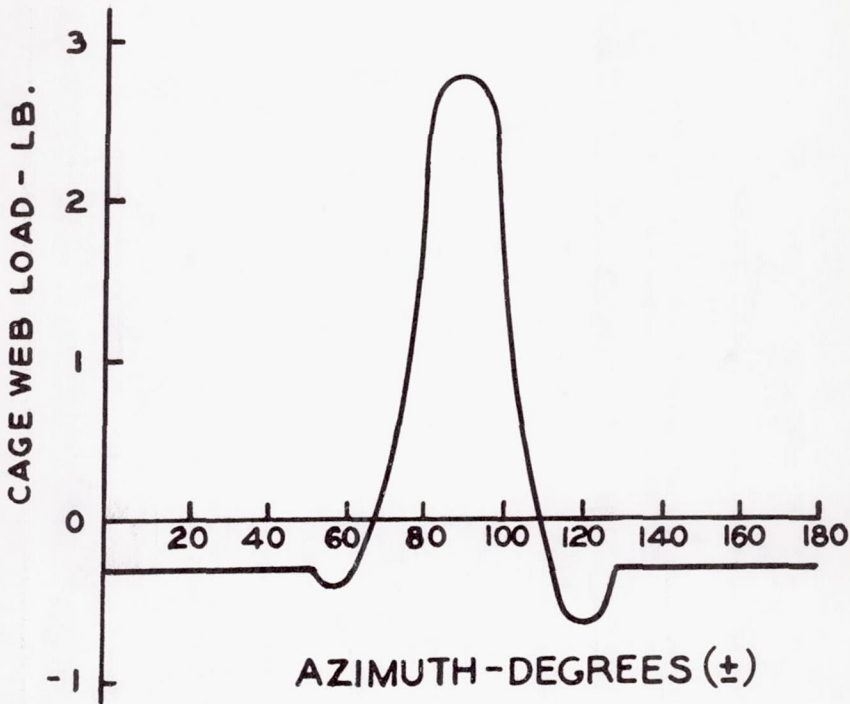


FIGURE 9.—Cage-to-roller load vs azimuth for a gas turbine main shaft cylindrical roller bearing. Thirty 12 mm  $\times$  12 mm rollers on a 6-in. pitch diameter. Roller ID/OD = 0.6, outer ring out-of-roundness = 0.01 in., radial load = 100 lb, shaft speed = 25 000 rpm.

forces in a normal force and moment equilibrium solution has demonstrated that:

- (1) Raceway control of ball motion does not occur in viscous-film lubricated ball bearings except in applications where ball-to-raceway skidding is present and then not as an absolute condition.
- (2) Ball rotation or slippage owing to gyroscopic moments must occur in a properly lubricated ball bearing.

Figure 10 indicates the ball loading considered in this analysis while figures 11 and 12 yield comparisons of analytical results against the test data of Shevchenko and Bolan (ref. 15) and Poplawski and Mauriello (ref. 16). Again possibly the non-Newtonian lubricant behavior accounts for the discrepancies between analytical and test data in the high-skid region. Figure 13 shows typical sliding velocity traverses along the major axes of the inner and outer contact ellipses in a skidding bearing.



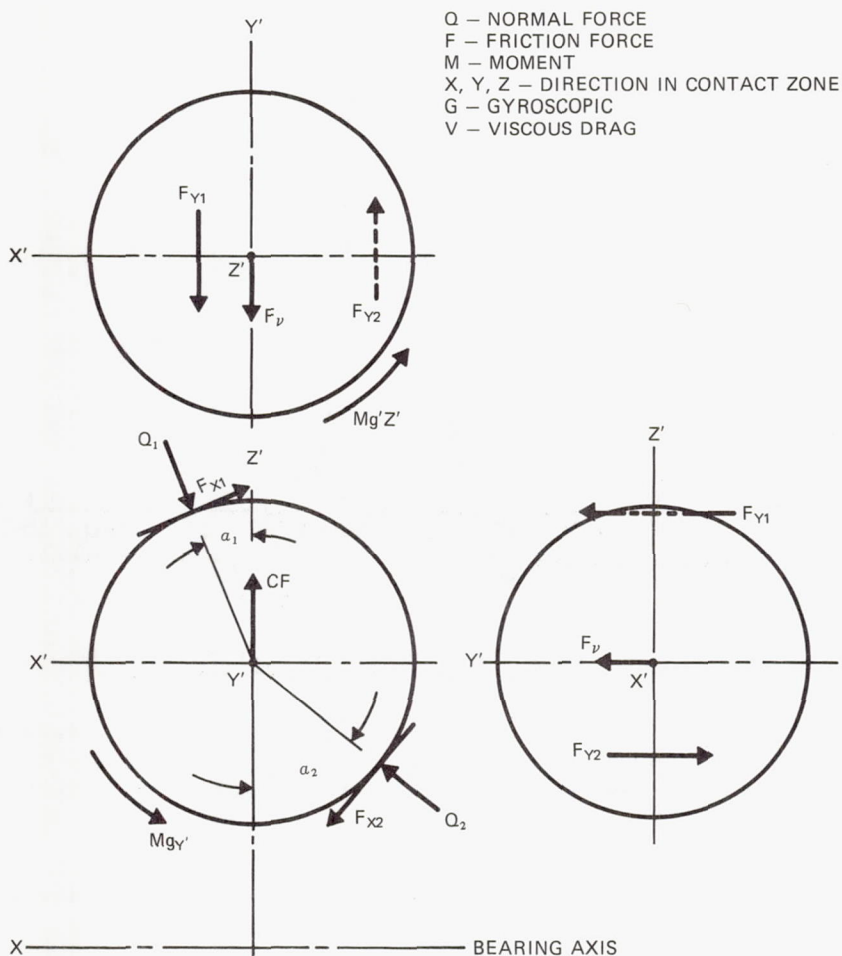


FIGURE 10.—Forces and moments acting on a ball.

As for the roller bearing, definition of the ball speeds permits a determination of cage-to-ball loading which in turn will effect the proper bearing and cage design in any application.

#### CONCLUSIONS

It has been demonstrated in the foregoing discussion that EHD lubricant film thickness and friction forces in the concentrated load contacts occurring in ball and roller bearings significantly affect:

- (1) Bearing endurance
- (2) Bearing friction torque
- (3) Bearing cage loading

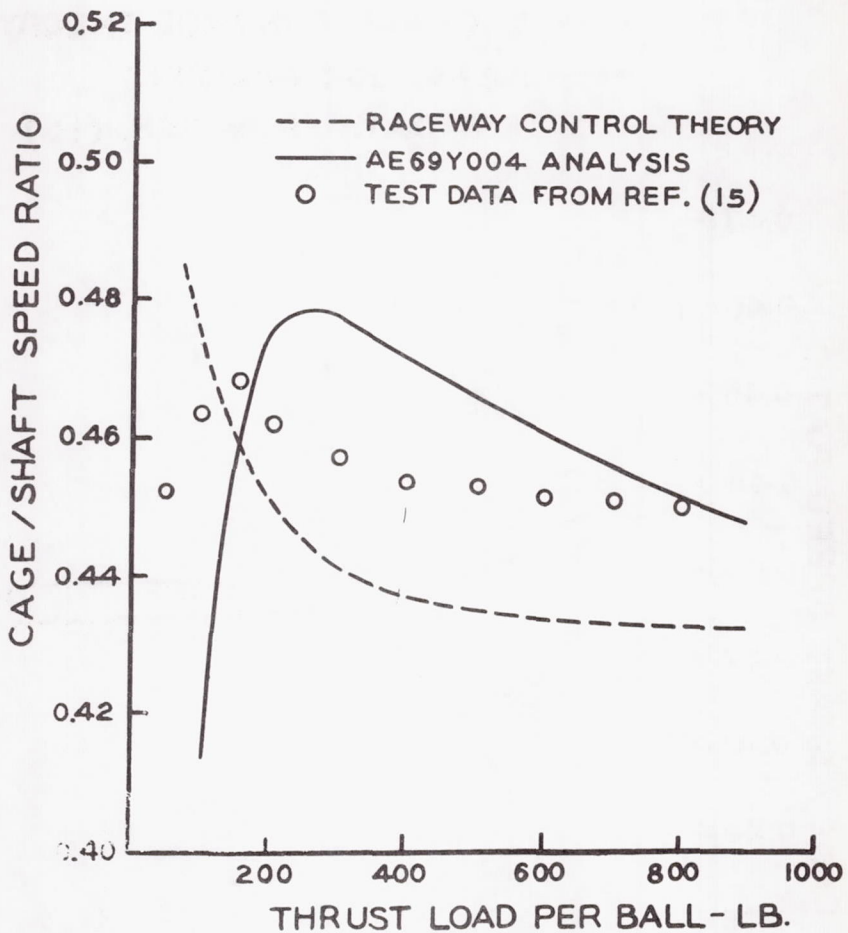


FIGURE 11.—Cage/shaft speed ratio vs thrust load per ball. 3 balls, 1.125 diameter, 9000 rpm.

Consequently proper bearing design for any viscous-film lubricated application depends upon a correct assessment of the lubricant performance in the bearing. To accomplish this, more information must be developed on:

- (1) Lubricant viscous properties as a function of pressure and temperature
- (2) Friction forces (or coefficients) in EHD lubricated contacts
- (3) Bearing fatigue life or endurance as a function of lubricant film thickness

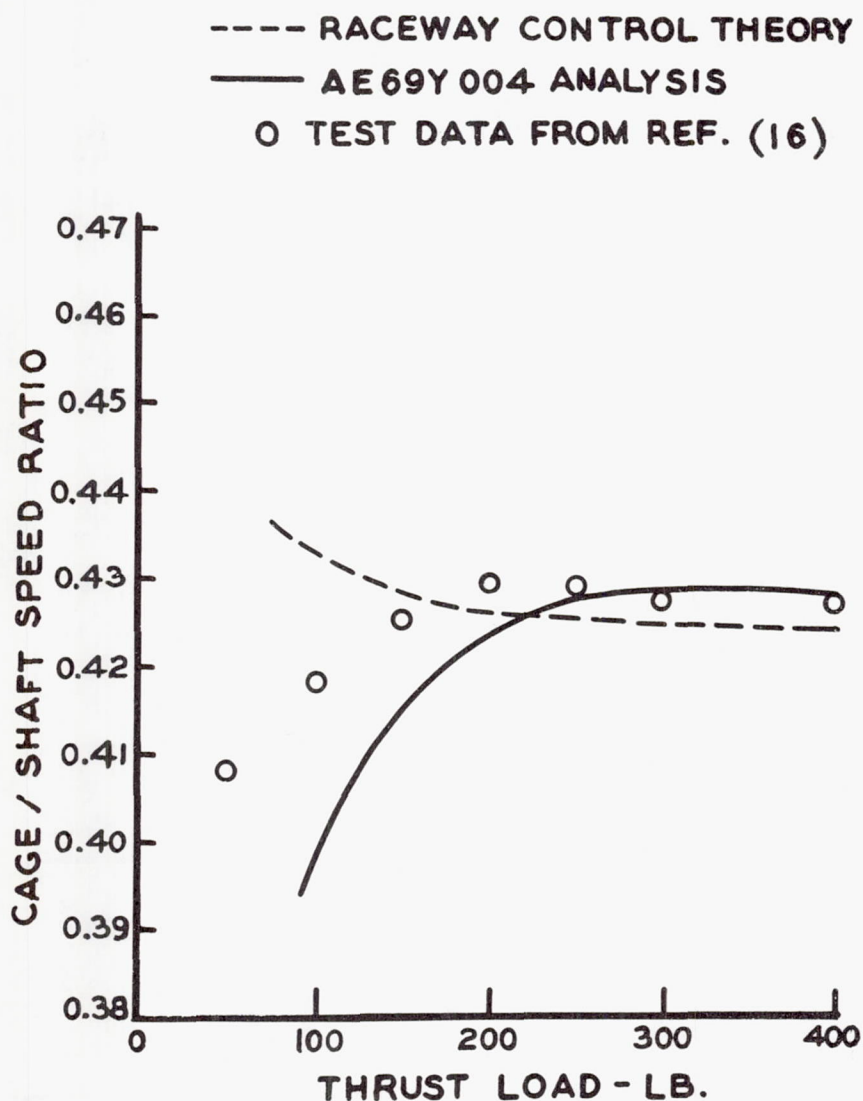


FIGURE 12.—Cage/shaft speed ratio vs thrust load. 35 mm  $\times$  62 mm ball bearing, 27 000 rpm.

#### DISCUSSION

P. M. Ku (Southwest Research Institute, San Antonio, Texas)

Messrs. Wellons and Harris have treated with admirable deftness the complex problems of rolling bearing fatigue and skidding, emphasizing the implications of elastohydrodynamics.



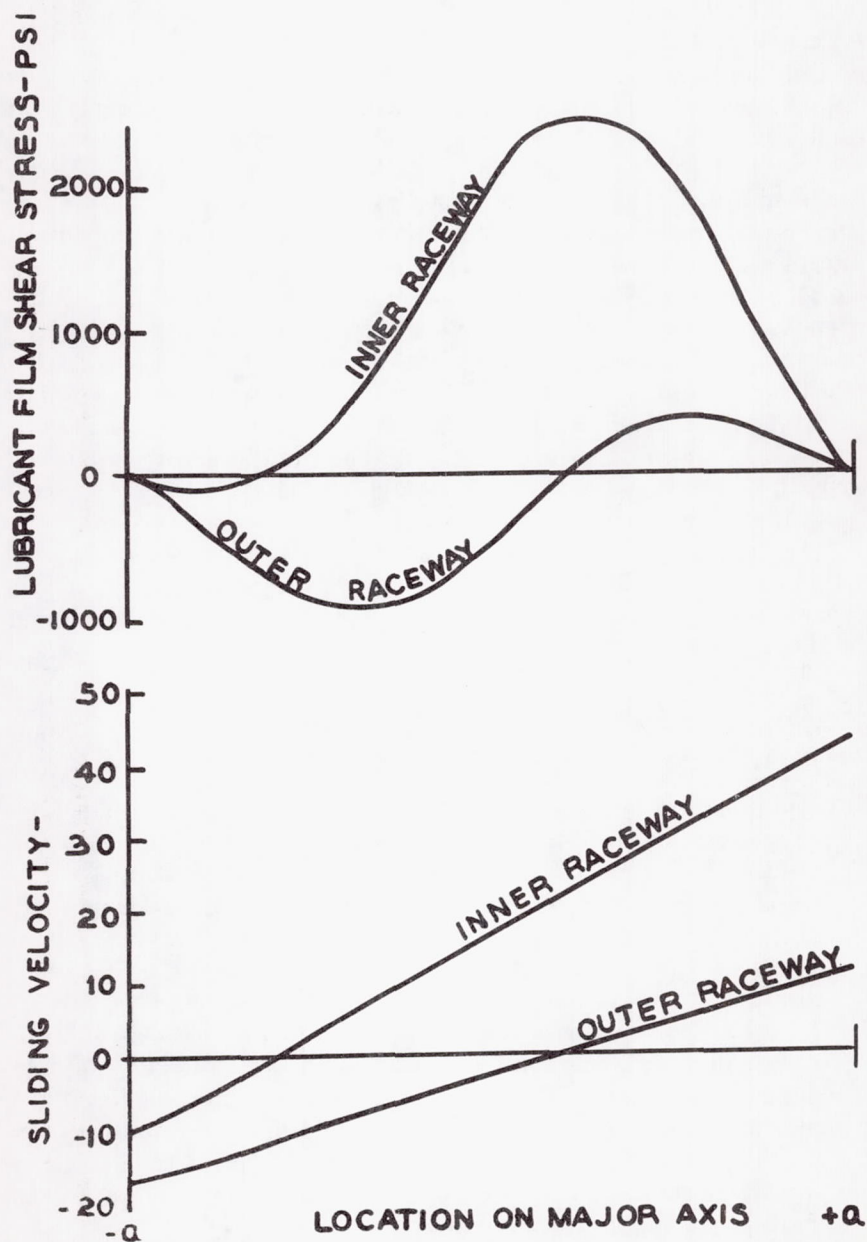


FIGURE 13.—Sliding velocity and lubricant shear stress vs location on the contact ellipse major axis. 140 mm  $\times$  220 mm angular-contact ball bearing, twenty 1-in. balls on a 7-in. pitch diameter, 30° free contact angle, 0.52 raceway grooves.

Rolling fatigue may be subsurface- or surface-oriented. The Lundberg-Palmgren life equation is based presumably on subsurface-oriented fatigue. Do the authors feel that this equation also applies to surface-oriented fatigue?

One would imagine that if fatigue occurs in the EHD regime and is subsurface-oriented, then only the rheological properties of the lubricant and the operating conditions would influence fatigue life, but not the chemical properties of the lubricant. On the other hand, for surface-oriented fatigue, the chemical properties of the lubricant might well be important. In figure 3, the authors show that the fatigue life increases steadily and uniquely with increasing  $\Lambda$  (the ratio of EHD film thickness to composite surface roughness). Does it mean that, in both elastohydrodynamic and boundary lubrication regimes, the fatigue is predominantly subsurface-oriented and therefore chemical effects are essentially absent? Further, the same  $\Lambda$  may be obtained by different combinations of lubricant viscosity, speed, load, and composite surface roughness. Does it mean that, as long as  $\Lambda$  is the same, the film pressure distribution and subsurface stress distribution are invariant regardless of the individual magnitudes of the several variables referred to?

The authors have discussed the use of elliptical outer race as a means of minimizing skidding. Their comments with regard to another possible approach, that of using several slightly oversized hollow rollers, would also be appreciated.

#### LECTURERS' CLOSURE

Mr. Ku's commentary is most welcome as always, and we will attempt to satisfy his queries categorically as follows:

1. The Lundberg-Palmgren theory is based upon contact fatigue failure emanating from material defects within a stressed volume near the surface. It encompasses a maximum orthogonal subsurface shear stress considered to be the cause of failure. This maximum shear stress occurs at fixed depth below the surface depending upon surface geometrics and contact load. According to the theory, the subsurface shear stress is a function of normal load only. It occurs to the writers, however, that if the surface shear stresses could be accommodated in conjunction with the normal stresses in the determination of subsurface shear stresses, the Lundberg-Palmgren theory might be logically extended to evaluate surface-oriented fatigue as a limiting case. It is known that as the surface shear stress increases the maximum subsurface shear stresses move closer to the stressed surface. In the limit, it must occur at the surface.
2. Considering the foregoing, the surface chemical effects can indeed affect subsurface fatigue life significantly. Chemical reaction of the contact surface with the lubricant or atmosphere tends to form (or

remove) surface films that substantially affect the coefficient of friction and subsequently the surface shear stress. Since the surface shear stress affects both the magnitude and depth of the maximum subsurface shear stress, it also affects the bearing fatigue life.

3. It seems to the writers that as the "lambda" ratio increases, so does that ratio of surface normal stress to surface shear stress owing to thicker lubricant films and decreased surface shear stress. Also the incidence of surface asperity contact lessens. Therefore, the depth of maximum subsurface shear stress tends to increase, and the magnitude tends to decrease.
4. It does not seem reasonable that the same "lambda," no matter how it is obtained, will always yield the same lubricant film pressure distribution and subsurface stress distribution. There exists a multiplicity of effects associated with a "lambda" ratio. The contemporary method of applying the "lambda" ratio is clearly simplistic, and there is no doubt that sometime in the future "lambda" will be subdivided into various complex components. Despite the foregoing, bearing industry technology today is far richer than before merely by the recognition of a lubrication-life parameter even though it now appears in simplistic format.
5. Roller bearing skidding is caused by insufficient normal loading between the bearing inner raceway and rollers. To prevent skidding, loading may be artificially induced (preload) between the rollers and raceways. Historically this has been accomplished with a flexible, elliptical outer ring which is sprung over the roller set. The minor axis of the outer ring ellipse is less than the diameter over the roller set, thus achieving selective roller preloading. Preload can also be accomplished with oversize rollers providing an interference fit-up. Solid rollers allowing for Hertzian deformations only are extremely sensitive to small amounts of interference and may tend to overload quickly and fail prematurely. Hollow rollers provide the required increased flexibility. On the negative side, however, anti-skid design, high-speed cylindrical roller bearings which employ hollow rollers must operate at sufficiently low levels of roller bending stress to preclude fatigue. Thus the design of these anti-skid bearings must be balanced between required minimal skidding properties and minimal roller bending stresses.

#### NOMENCLATURE

$a_1$	reliability-life factor
$a_2$	lubrication-life factor
$a_3$	material-life factor
$C$	basic load rating, lb
$D$	ball or roller diameter, in.



$f_c$	material and geometry factor
$F_a$	applied axial load, lb
$F_r$	applied radial load, lb
$H$	lubrication $H$ -factor relating to geometry
$i$	number of bearing rows
$L$	fatigue life corresponding to reliability $a_1$ , millions of revolutions
$L_{10}$	rating life, millions of revolutions
$N$	speed, rpm
$P$	equivalent radial load, lb
$P_o$	static equivalent load, lb
$u$	surface velocity, in/sec
$V$	rotation factor
$X$	radial load factor
$Y$	axial load factor
$z$	distance in the $z$ -direction, in.
$Z$	number of balls or rollers
$\alpha$	contact angle, degrees; or pressure coefficient of viscosity, $\text{psi}^{-1}$
$\Lambda$	lambda ratio
$\mu_o$	dynamic viscosity at atmospheric pressure, $\text{lb-sec/in}^2$
$\mu$	dynamic viscosity, $\text{lb-sec/in}^2$
$\tau$	lubricant shear stress, psi

## REFERENCES

1. ANON.: AFBMA Standard, Section 9: Method of Evaluating Load Ratings for Ball Bearings. Revision 4, 1960.
2. ANON.: AFBMA Standard, Section 11: Method of Evaluating Load Ratings for Roller Bearings.
3. LUNDBERG, G.; AND PALMGREN, A.: Dynamic Capacity of Rolling Bearings. Acta Polytechnica, Mech. Engrg. Series 1, Roy. Swedish Acad. Engrg. Sci., no. 3, 1947, p. 3.
4. LUNDBERG, G.; AND PALMGREN, A.: Dynamic Capacity of Roller Bearings. Acta Polytechnica, Mech. Engrg. Series 2, Roy. Swedish Acad. Engrg. Sci., 1952, p. 96.
5. GRUBIN, A. N.; AND VINOGRADOVA, I. E.: Investigation of the Contact of Machine Components. Moscow, Ts NITTMASh, Book 30, 1949 (D.S.I.R., London, Translation 337).
6. SIBLEY, L. B.; AND ORCUTT, F. K.: Elastohydrodynamic Lubrication of Rolling Contact Surfaces. ASLE Trans., vol. 4, 1961, pp. 234-249.
7. HARRIS, T. A.: A New Way to Evaluate Bearing Lubrication. Product Engineering, April 12, 1965, pp. 76-81.
8. TALLIAN, T. E., et al.: Elastohydrodynamic Film Effects on the Load-Life Behavior of Rolling Contacts. ASME paper 65-Lubs-11, 1965.
9. HARRIS, T. A.: Predicting Bearing Reliability. Machine Design, January 3, 1963, pp. 129-132.
10. HARRIS, T. A.: An Analytical Method to Predict Skidding in High Speed Roller Bearings. ASLE Trans., vol. 9, 1966, pp. 229-441.
11. HARRIS, T. A.; AND AARONSON, S.: An Analytical Investigation of Cylindrical Roller Bearings Having Annular Rollers. ASLE Preprint 66LC-26.

12. HARRIS, T. A.; AND AARONSON, S.: An Analytical Investigation of Skidding in a High Speed Cylindrical Roller Bearing Having Circumferentially-Spaced, Pre-loaded Annular Rollers. *Lub. Engrg.*, vol. 24, 1968, pp. 30-34.
13. JOHNSON, K. L.; AND CAMERON, R.: Shear Behavior of Elastohydrodynamic Oil Films at High Rolling Contact Pressures. *Proc. Inst. Mech. Engrs.*, vol. 182, pt. 1, 1967-68.
14. CHENG, H. S.: Refined Solution to Thermal-Elastohydrodynamic Lubrication of Rolling and Sliding Cylinders. *ASLE Trans.*, vol. 8, 1965, pp. 397-410.
15. SHEVCHENKO, R. P.; AND BOLAN, P.: Visual Study of Ball Motion in a High Speed Thrust Bearing. *SAE Paper No. 37*, January 14-18, 1957.
16. POPLAWSKI, J.; AND MAURIELLO, J.: Skidding in Lightly Loaded High Speed Ball Thrust Bearings. Presented at the ASME Spring Lubrication Symposium, San Francisco, June 19, 1969.

**Page intentionally left blank**



# Gear Design Considerations

**W. COLEMAN**

**Gleason Works  
Rochester, New York**

This paper discusses the parameters of gears and gear lubrication that affect wear, pitting, and scoring. Formulas for sliding and rolling velocities and for relative radii of profile and lengthwise curvatures between gear teeth are listed. The effects of load sharing between teeth are considered, and formulas for contact stress are developed. The duration of tooth contact and the variation in coefficient of friction are considered. The aspects of oil film thickness and critical temperatures are presented in relation to gear tooth wear and scoring. A discussion of gear testing and simulated gear testing by means of rollers emphasizes the advantages of each.

**T**OOOTH SURFACES have experienced distress ever since gearing was first introduced. In the early years empirical formulas were devised for selecting gears of adequate size to minimize or eliminate the likelihood of surface distress. This generally resulted in grossly over-sized gears. With the advent of worm gears and hypoid gears, the problems of gear design became more serious and the need for improved lubricants became obvious. During the late 1920's and early 1930's, EP lubricants were developed explicitly for use in automotive hypoid rear axles. Without such lubricants it is doubtful whether the hypoid gear would have been successful.

Up to the time of Wilfred Lewis (ref. 1), very little science was applied to gear design. As pointed out by Lewis, the quality of most gears was so poor that modern formulas would have had little meaning. With the advent of the automotive industry, the need for more accurate and quiet gears became obvious. During the first quarter of the present century, great strides were made in the improvement of gear quality and methods of gear manufacture. Heat treating of gears was commonplace, and the designer was no longer limited by quality and strength.

Because of their economic advantage, many gears of lower hardness continued to be made, and wear was a common problem. As the loads increased, pitting of the tooth surfaces became a more common problem.

And with the advent of high speed aircraft gearing in the 1930's, scoring began to enter the picture.

All of the above have led to interest in obtaining better formulas, improved materials, higher quality manufacturing methods, and new lubricants in order to reduce surface failures and increase the load-carrying capacity of gears.

#### TYPES OF GEAR SURFACE FAILURES

Many types of surface failures have been named and defined for gears; but only the major ones will be listed and defined here. For heavily-loaded low-speed gears five phenomena may occur:

- (1) adhesive wear of the tooth surfaces due to high contact stress

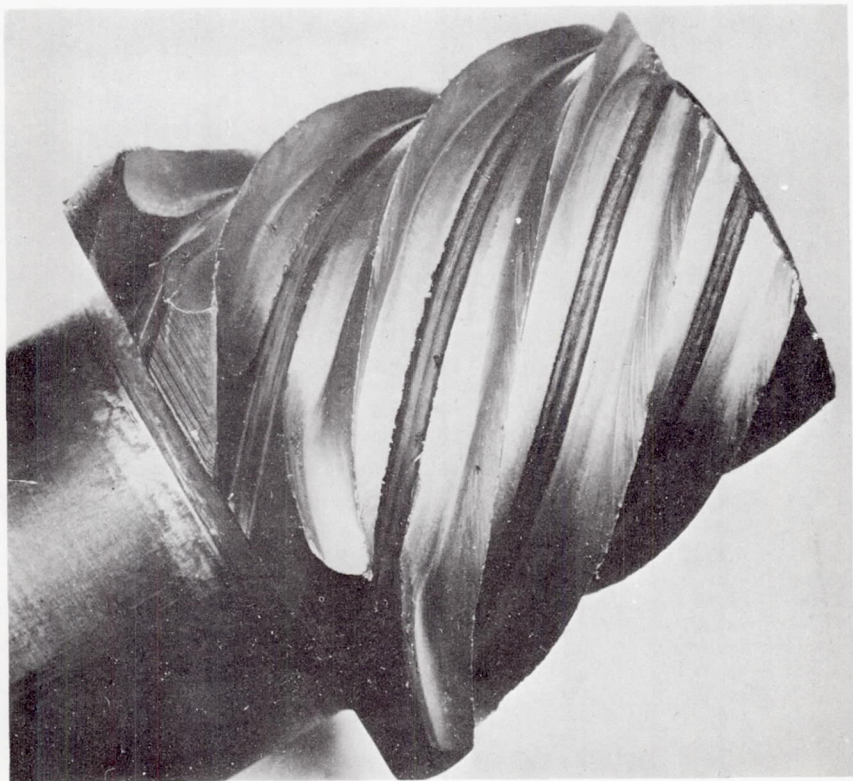


FIGURE 1.—Example of extreme wear on the tooth surfaces of a hypoid pinion. In this case the teeth have worn badly on both sides until there is nearly no tooth left. The fact that the tooth did not ultimately break indicates that the tooth loads were light. The initial failure was caused by a breakdown of the oil film.

combined with low-speed sliding, especially on unhardened or through-hardened gears (fig. 1);

(2) pitting of the tooth surface due to a high contact pressure at the surface (fig. 2);

(3) case crushing of the tooth surface on case-hardened gears due to high contact pressure combined with inadequate heat treatment (fig. 3);

(4) ridging of the tooth surface resulting from plastic flow of the surface and subsurface layers as a result of surface pressure and sliding (fig. 4);

(5) rippling of the tooth surface, which is also a plastic flow of the surface layer and can ultimately lead to rapid wear (figs. 5 and 6).

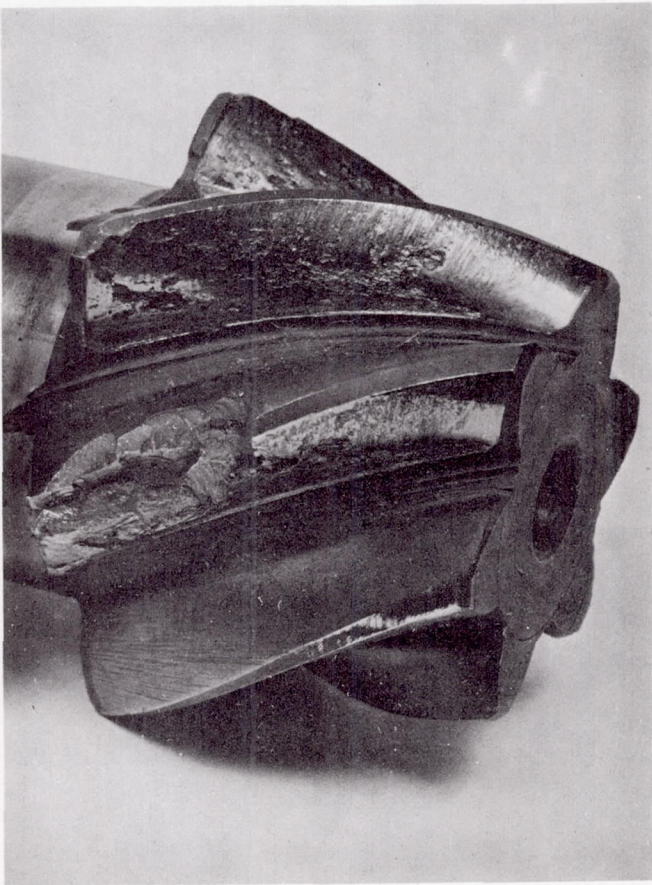


FIGURE 2.—Example of pitting on a case-hardened tooth surface. This usually occurs at or near the point of highest contact pressure in a region where little sliding is present. Note that this type of failure is characterized by cracks at an oblique angle to the surface.





FIGURE 3.—Case crushing on a case-hardened gear tooth. Note that the direction of the cracks is perpendicular to the surface and that failure occurs below the surface near the boundary between the case and core. It is usually the result of insufficient material strength at the point of maximum subsurface stress.

In all of these cases, surface pressure is a prime factor contributing to failure. In addition, all except case crushing are influenced by lubrication.

Scoring results from a combination of surface pressure and a high sliding velocity. For moderately-loaded high-speed gears, this is the most likely type of surface failure (fig. 7).

Two additional types of surface failure are: (1) abrasive wear due to contamination of the lubricant with metallic particles, and (2) corrosive wear due to water or other substances which attack the gear material chemically. Abrasive wear and adhesive wear have a similar appearance and are difficult to differentiate. Generally adhesive wear is influenced by metallic particles from the gear tooth surfaces only, whereas abrasive wear results from contamination by metallic or nonmetallic particles not originating from the gear tooth surfaces.

Since abrasive wear, corrosive wear, and case crushing are not generally



FIGURE 4.—Ridging is a common occurrence on hypoid gear teeth. The ridges are in the direction of sliding as in the case of scoring, but are usually deeper. Microscopic examination of the thin layer just below the surface will confirm that this is a plastic deformation.

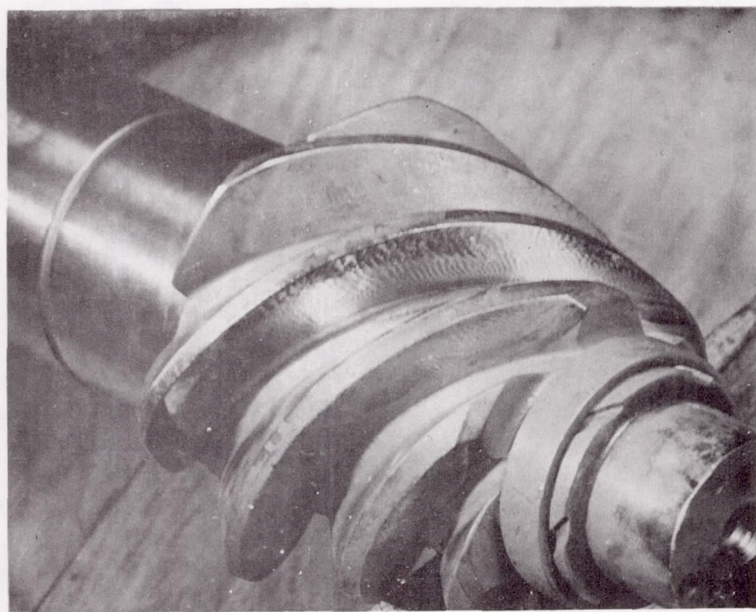


FIGURE 5.—Rippling is also common on hypoid pinion teeth. It also results from plastic flow of the surface layer and can often lead to rapid wear.

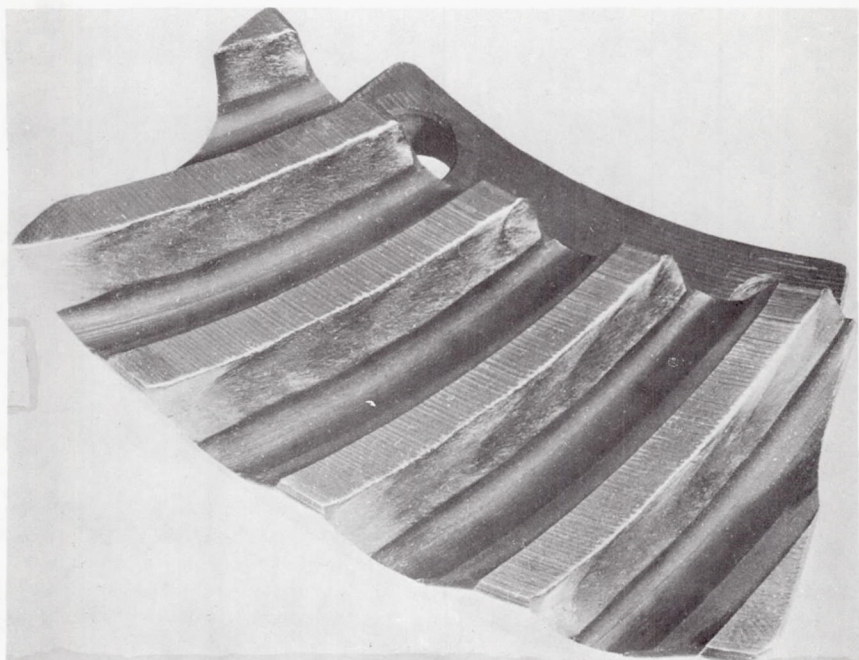


FIGURE 6.—This illustration shows a combination of rippling and wear.

associated with lubrication, these types will not be considered further in this paper.

Figure 8, which is taken from Dudley (ref. 2) and has been modified as suggested by Wellauer,\* shows graphically the various regimes of surface distress.

#### PURPOSE OF LUBRICATION

Lubrication serves two major purposes in gear operation—to prevent metal-to-metal contact and to cool the gears. It is obvious to most engineers that metal-to-metal contact will result in rapid wear, a relatively high coefficient of friction, and possible scoring. However, it is not always recognized that the cooling effects of the lubricant are equally important. The high temperatures, which are generated at the point of contact by the combination of pressure and sliding, are dissipated by the heat flow through the teeth to the gear blank. But the blank temperature must be maintained at a sufficiently low value that the heat transfer from the point of contact will continue at a rapid and steady rate. This cannot be accomplished unless there is a sufficient rate of cooling of the gear blank.

---

\* E. J. Wellauer's suggestions were made in an unpublished discussion of the Role of Lubrication in Gear Design, Univ. of Wisconsin, November 1968.





FIGURE 7.—Scoring is a welding and subsequent tearing apart of the tooth surfaces due to a breakdown of the oil film.

#### GEAR TOOTH EFFICIENCY

No mechanical element has an efficiency of 100 percent. It is difficult to measure the true efficiency of gear teeth since the gears must be mounted in bearings in order to function, and the measured efficiency is that of the gears and bearings. However, many gear meshes including the bearings have a very high efficiency—above 99 percent. With crossed-axis gearing such as crossed-helical gears, hypoid gears, and worm gears the efficiency may be considerably less due to the higher amount of sliding between the teeth.

The actual efficiency in any gear mesh is a function of the particular arrangement of the gear axes, the gear ratio, the tooth design, and the load on the gears. In general, the greater the offset between the shafts on crossed-axis gearing, the lower will be the efficiency. With high gear ratios the efficiency will be less. Tooth design should be based on the minimum amount of sliding if maximum efficiency is to be achieved. And finally the efficiency decreases as the load decreases. Therefore, low efficiency means a high rate of heat generation.

High efficiency means less heat generated; and when less heat is generated, less cooling is required. In order to more clearly understand the elements of gear design which affect lubrication most critically, it is necessary to analyze the nature of tooth contact.

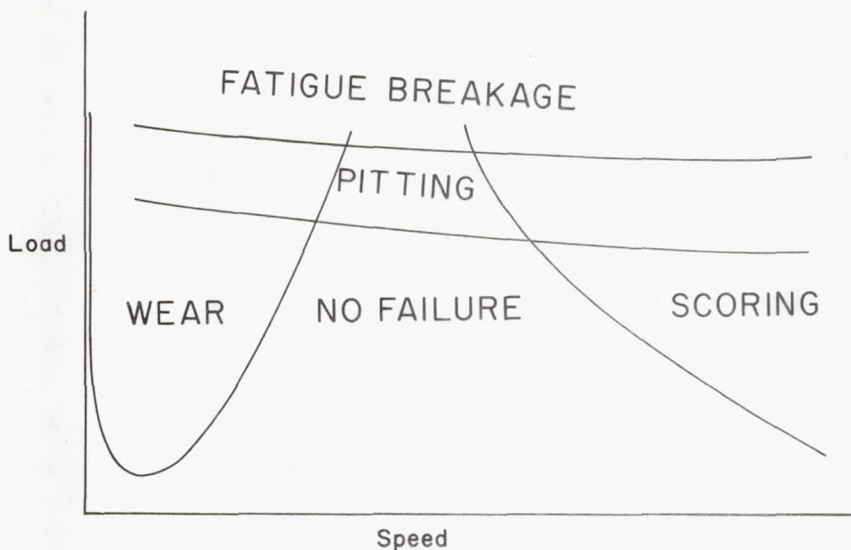


FIGURE 8.—Graph showing the various failure regimes encountered in gear teeth. For carburized gears the pitting and fatigue breakage areas may be reversed.

#### NATURE OF TOOTH CONTACT

Many spur gear teeth are manufactured with a full-length tooth contact, that is, with an infinite lengthwise radius of curvature on mating teeth. Therefore, the only curvature is the tooth profile curvature. For involute tooth profiles the radius of profile curvature is a simple value to calculate. The instantaneous tooth contact pattern will be a straight line extending from one end of the tooth to the other if there is no elastic deformation of the teeth, or will be in the form of a long narrow rectangle extending from one end of the tooth to the other when elastic deformation is present. When the teeth are crowned, lengthwise curvature of the teeth is introduced which results in point contact if there is no elastic deformation, or an elliptical tooth contact pattern when elastic deformation is present. This rectangle or ellipse will move up and down the tooth profile as the gears rotate (fig. 9).

Helical gears have an instantaneous contact similar to that described above for spur gears, except that the rectangle or ellipse of instantaneous contact is at an oblique angle to the direction of the tooth.

Bevel and hypoid gears are nearly always produced with lengthwise crowning of the teeth. Therefore, the instantaneous tooth contact pattern will be a theoretical point if there is no elastic deformation, but will be an ellipse as shown in figure 10 when elastic deformation is present. Point O is a point of contact. Line QR is the major axis of the contact

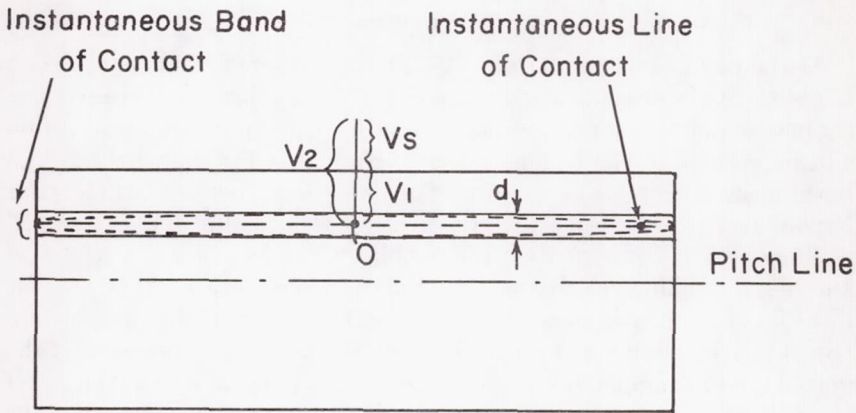


FIGURE 9.—Diagram showing an instantaneous rectangular band of contact (non-crowned teeth) and ellipse of contact (crowned teeth) for a typical spur gear. Shown also is the theoretical line of contact and the velocity vectors at point O.

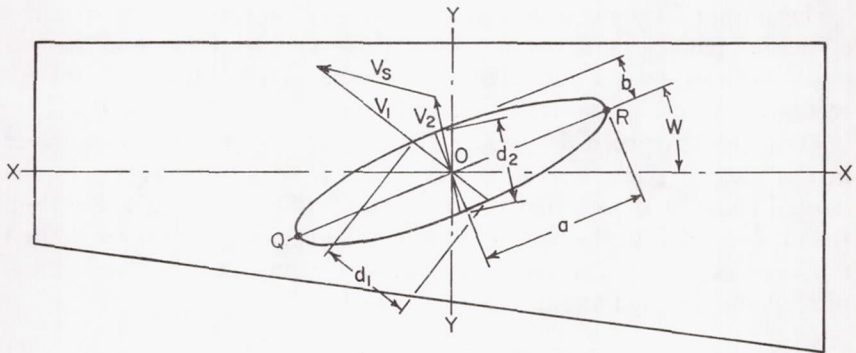


FIGURE 10.—This diagram shows a typical point of contact O on a hypoid gear tooth in the tangent plane. Shown also are the resultant velocity vectors, the instantaneous area (ellipse) of contact under load, and the distances of traverse across the ellipse of contact. This is the most general case of contact between gear teeth.

ellipse under a given load and at a particular instant of time during the gear rotation. Because of the oblique direction of the teeth on hypoid gears, the line of contact is inclined to the lengthwise direction of the tooth. The inclination is specified by the angle,  $w$ , on the diagram. As the gears rotate, the theoretical point of contact progresses across the tooth along a "path of contact" in a direction usually up and down the tooth profile (direction YY). This is essentially identical to the conditions in helical gear teeth with lengthwise crowning.



## Velocity Vectors

At the point of contact  $O$  (figs. 9 and 10), vectors  $v_1$  and  $v_2$  are drawn to represent the directions and magnitudes of the velocities of the point of contact on pinion and gear, respectively. Vector  $v_s$  represents the direction and magnitude of the sliding velocity. Note that for spur, helical, and bevel gears, the velocity vectors are as shown in figure 9; whereas for hypoid gears, the velocity vectors are as shown in figure 10.

Since heat is generated by the rubbing or sliding of one surface on another, the sliding velocity between the two tooth surfaces is one obvious measure of the heat generated at any instant. However, the temperature rise at a given point is not a function only of the sliding component of the velocity. For example, if one wishes to file the point of a nail, the tip of the nail will become hot after a few strokes, whereas the surface of the file remains cool. The sliding velocity between the two members is the same since this is a relative velocity. However, the point of contact on the nail remains stationary, whereas the point of contact on the file moves along the surface of the file.

On spur, helical, and bevel gears, the point of contact remains constant on neither member, and the directions of the velocity vectors on the two tooth surfaces are the same. On crossed-axis gearing, however, the directions of the velocity vectors are different.

Formulas have been derived by the author (ref. 3) for the velocities in the tangent plane for the general case, as seen in figure 11. In the normal plane, it is seen that the distance along the path of action from the pitch point  $P$  to the point  $Q$  at which the velocity vectors are desired is  $z_o$ , see figure 12. The tangential velocity vector in this plane at the pitch point is  $v_n$  and at point  $Q$  is  $v_n'$ . Then

$$v_n = \omega_P (A \sin \gamma) \cos \psi \quad (1)$$

$$v_n' = \omega_P z_o \sin \phi_n \cos \gamma \cos \psi + v_n \quad (2)$$

At point  $Q$  in the tangent plane, the velocity vector in the lengthwise direction on the pinion tooth is  $v_{FP}$  and the velocity vector in the profile direction is  $v_{pP}$ . The resultant absolute velocity vector on the pinion is  $v_1$ . Then

$$v_{FP} = v_n \left[ \tan \psi + |z_o| \sin \phi_n \left( \frac{\tan \psi}{A \tan \gamma} \right) - |z_o| \cos \phi_n \left( \frac{1}{A \cos \psi} \right) \right] \quad (3)$$

$$v_{pP} = v_n \left[ \sin \phi_n + z_o \left( \frac{1}{A \tan \gamma} \right) \right] \quad (4)$$

$$v_1 = \sqrt{v_{FP}^2 + v_{pP}^2} \quad (5)$$

In like manner, the equations for the mating gear tooth are

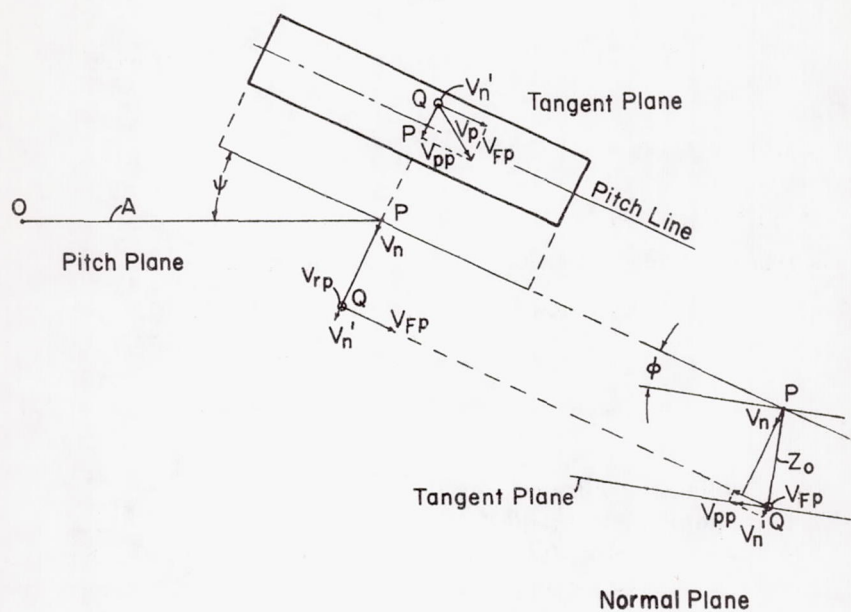


FIGURE 11.—View of a pinion tooth showing velocity vectors in pitch plane, normal plane, and tangent plane.

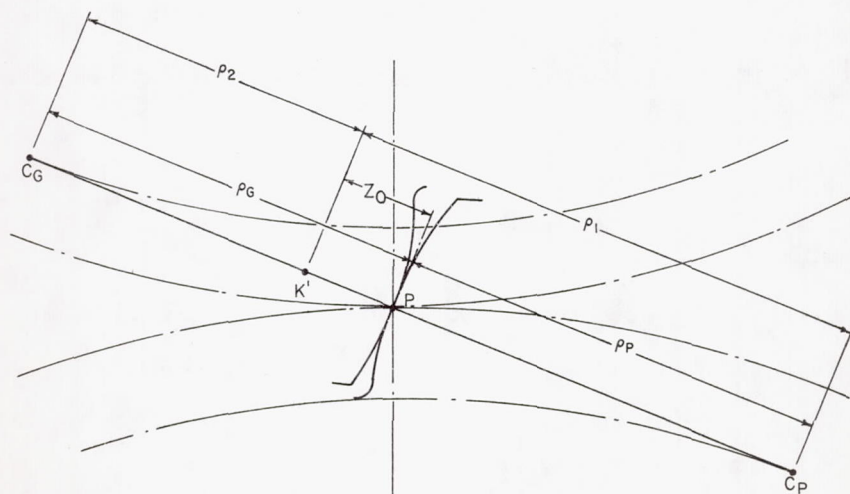


FIGURE 12.—Profile radii of curvature at pitch point,  $P$ , and at point of load concentration,  $K'$ .

$$v_{FG} = v_n \left[ \tan \psi - |z_o| \sin \phi_n \left( \frac{\tan \psi}{A \tan \Gamma} \right) - |z_o| \cos \phi_n \left( \frac{1}{A \cos \psi} \right) \right] \quad (6)$$

$$v_{pG} = v_n \left[ \sin \phi_n - z_o \left( \frac{1}{A \tan \Gamma} \right) \right] \quad (7)$$

$$v_2 = \sqrt{v_{FG}^2 + v_{pG}^2} \quad (8)$$

The relative sliding velocities may now be determined:

$$v_F = v_{FP} - v_{FG} = \text{lengthwise sliding velocity.} \quad (9)$$

$$v_p = v_{pP} - v_{pG} = \text{profile sliding velocity.} \quad (10)$$

$$v_s = \sqrt{v_F^2 + v_p^2} = \text{resultant sliding velocity.} \quad (11)$$

where

$\omega_P$  = pinion angular velocity.

$\phi_n$  = normal pressure angle.

$A$  = mean cone distance for bevel gears. On hypoid gears the mean cone distance on pinion and gear are not equal. Use the value for the corresponding member in the foregoing formulas.  $A$  will be infinite for spur and helical and crossed-helical gears.

$\gamma, \Gamma$  = pitch angles of pinion and gear, respectively. Use zero for spur and helical and crossed-helical gears.

$\psi$  = spiral angle or helix angle. On hypoid gears and crossed-helical gears, the spiral angle or helix angle on pinion and gear are not equal. Use the value for the corresponding member in the foregoing formulas.  $\psi$  will be zero for spur gears and straight bevel gears.

For spur and helical gears, the following substitutions should be made:

$$A \sin \gamma = r = \text{pinion pitch radius}$$

$$A \tan \gamma = r = \text{pinion pitch radius}$$

$$A \tan \Gamma = R = \text{gear pitch radius}$$

$$\frac{1}{A \cos \psi} = 0$$

The above formulas therefore reduce to the following for helical gears:

$$v_n = \omega_P r \cos \psi \quad (1a)$$

$$v_n' = \omega_P z_o \sin \phi_n \cos \psi + v_n \quad (2a)$$

$$v_{FP} = v_n \left[ \tan \psi + |z_o| \sin \phi_n \left( \frac{\tan \psi}{r} \right) \right] \quad (3a)$$



$$v_{pP} = v_n \left[ \sin \phi_n + \frac{z_o}{r} \right] \quad (4a)$$

$$v_1 = \sqrt{v_{pP}^2 + v_{pP}^2} \quad (5a)$$

$$v_{FG} = v_n \left[ \tan \psi - |z_o| \sin \phi_n \left( \frac{\tan \psi}{R} \right) \right] \quad (6a)$$

$$v_{pG} = v_n \left[ \sin \phi_n - \frac{z_o}{R} \right] \quad (7a)$$

$$v_2 = \sqrt{v_{FG}^2 + v_{pG}^2} \quad (8a)$$

$$v_F = v_n |z_o| \sin \phi_n \tan \psi \left( \frac{1}{r} + \frac{1}{R} \right) \text{ lengthwise sliding velocity} \quad (9a)$$

$$v_p = v_n z_o \left( \frac{1}{r} + \frac{1}{R} \right) \quad \text{profile sliding velocity} \quad (10a)$$

$$v_s = v_n |z_o| \left( \frac{1}{r} + \frac{1}{R} \right) \frac{\cos \psi_b}{\cos \psi} \quad \text{resultant sliding velocity} \quad (11a)$$

where  $\psi_b$  = base helix angle

It is generally recognized that helical gears have no lengthwise sliding in the axial direction. However, it will be noted that helical gears do have both lengthwise sliding (eq. 9a) and profile sliding (eq. 10a) in the tangent plane at every point except the pitch point ( $z_o$  equals zero). The magnitude of the lengthwise sliding is, of course, very small.

#### Tooth Sliding

Although much attention is given to the control of sliding velocities and contact pressures between gear teeth, not much thought is given to the actual nature of tooth sliding. When the teeth first engage, the tip of the driven member engages the flank of the driving member with a scraping action during the arc of approach, which under certain circumstances can scrape the oil film off the tooth surface of the driving member. This can be reduced or eliminated by modifying the tip of the driven member so that it will not contact at its tip. The provision for tip relief is quite common on spur gears today, but it may not always be present. Bevel gears are usually produced with the equivalent of tip relief.

On gears where tip relief is not present or where the relief is not sufficiently great, there is a potential danger of scoring or gouging of the flank on the driving member. This is an effect that should be considered on any highly loaded gears.

The design of gears with all recess action (all addendum on the driving member) avoids this danger of tip scraping, but will result in gears with a higher sliding velocity at the tip of the driving member which may cause the gears to be more vulnerable to scoring.

#### Relative Radius of Profile Curvature

From figure 12 it can be seen that at point  $K'$  the relative radius of profile curvature  $\rho_o$  of the two contacting tooth profiles is

$$\rho_o = \frac{\rho_1 \rho_2}{\rho_1 + \rho_2} \quad (12)$$

where  $\rho_1 = \rho_P + z_o$  = radius of profile curvature on the pinion.

$\rho_2 = \rho_G - z_o$  = radius of profile curvature on the gear.

$\rho_P$  = radius of profile curvature at pitch point on pinion in the mean normal section.

$\rho_G$  = radius of profile curvature at pitch point on gear in the mean normal section.

#### Relative Radius of Lengthwise Tooth Curvature

For spur and helical gears with uncrowned teeth, the relative radius of lengthwise curvature will be infinite. For gear teeth that are crowned in the lengthwise direction, the relative radius of lengthwise curvature  $\rho_L$  will be

$$\rho_L = \frac{F^2}{8\Delta \cos^2 \psi} \quad (13)$$

where  $F$  = face width.

$\Delta$  = crowning between teeth. This is the total separation between the pinion and gear tooth surfaces at the ends of the teeth.

$\psi$  = helix angle.

Formulas have been derived by Baxter (ref. 4) that will analyze the contact between two mating hypoid gear teeth based on the specific method of generation actually employed. However, these formulas do not consider tooth bending under load. Approximate formulas have been developed by the author (ref. 3) based on the assumption that the tooth contact under load will remain within the boundaries of an ellipse that is tangent to the extremities of the tooth boundary as shown in figure 13. This assumption is based on teeth that are crowned in the lengthwise direction. It has the advantage of simplifying the calculations by defining the tooth boundaries with the equation of an ellipse rather than with the discontinuous function for the equation of a rectangle.

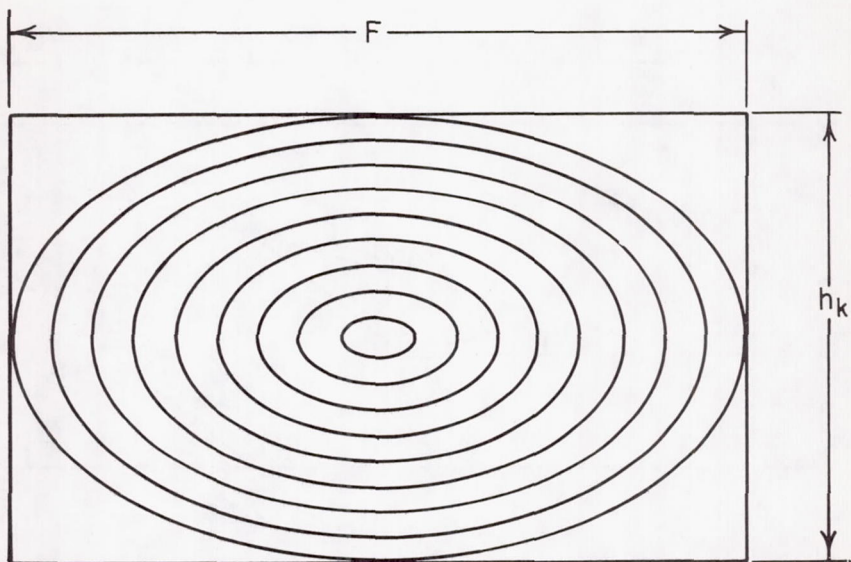


FIGURE 13.—Tooth-contact contour lines under full load.

#### Length of the Line of Contact

The length of the line of contact on an uncrowned spur gear tooth will be equal to the face width  $F$ . The length of the line of contact  $s_G$  on an uncrowned helical gear tooth is dependent upon the gear face width  $F$ , the base helix angle  $\psi_b$ , the length of action  $Z$ , and the position of the line of contact on the tooth. Figure 14 shows the four possible cases. These are represented below by the corresponding four equations:

$$\text{Case I} \quad s_G = \frac{K}{\sin \psi_b} \quad \text{when } K < Z \text{ and } K \leq F \tan \psi_b \quad (14)$$

$$\text{Case II} \quad s_G = \frac{F}{\cos \psi_b} \quad \text{when } K < Z \text{ and } K > F \tan \psi_b \quad (15)$$

$$\text{Case III} \quad s_G = \frac{Z}{\sin \psi_b} \quad \text{when } K > Z \text{ and } K < F \tan \psi_b \quad (16)$$

$$\text{Case IV} \quad s_G = \frac{Z - K + F \tan \psi_b}{\sin \psi_b} \quad \text{when } K > Z \text{ and } K > F \tan \psi_b \quad (17)$$

where  $K = p_b(N' - 1)$ .



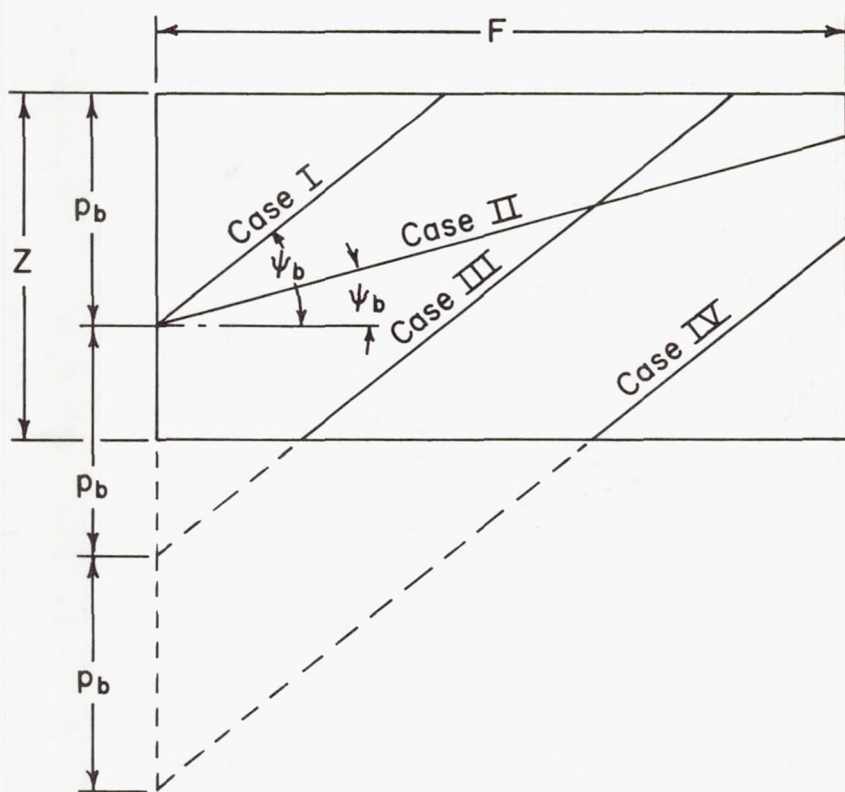


FIGURE 14.—Lines of contact on a noncrowned helical gear tooth. Cases I, III, and IV are for the same base helix angle.

$p_b$  = base pitch.

$N'$  = assumed integer between 1 and  $N_L$ .

$N_L = \frac{F \sin \psi_b + Z \cos \psi_b}{p_N} + 1$  to next lower integer.  $N_L$  is the minimum

number of lines of contact.

$p_N$  = normal base pitch.

For teeth that are crowned in the lengthwise direction, the length of the line of contact through point  $K$  (fig. 15) is the chord  $VW$  as derived by the author (ref. 3):

$$S_G = \frac{FZ\eta_1}{\eta^2} \quad (18)$$

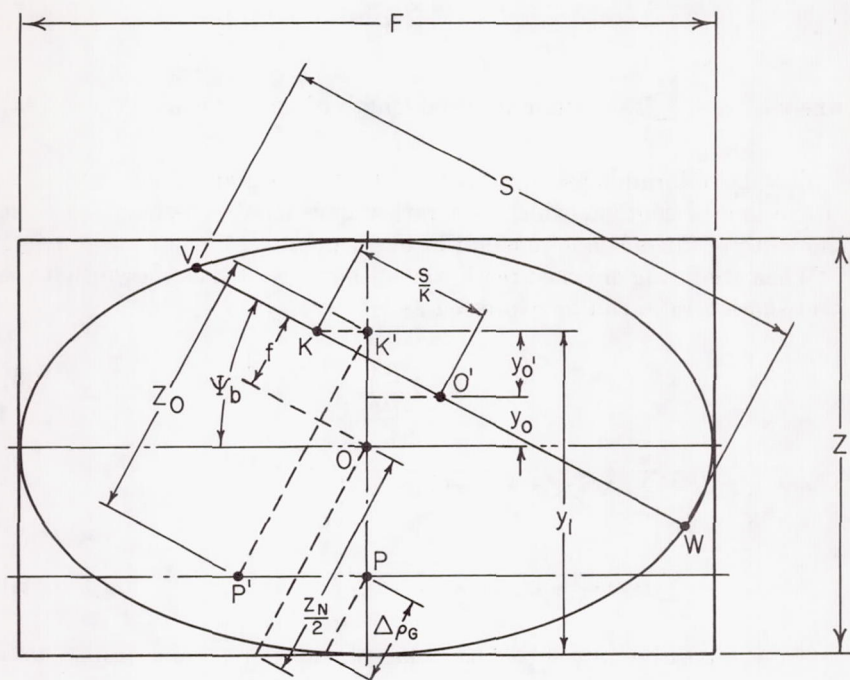


FIGURE 15.—Diagram showing chord of ellipse through point K.

$$\text{where } \eta_1 = \sqrt{p_N^2 m_o^2 + 4f^2}. \quad (18a)$$

$p_N$  = mean normal base pitch.

$m_o = \sqrt{m_F^2 + m_p^2}$  = modified contact ratio (contact ratio within boundaries of ellipse of tooth bearing in zone of action).

$m_F$  = face contact ratio.

$m_p$  = transverse (profile) contact ratio.

$f$  = distance from center of surface of action O, to line of contact, VW, measured in the normal direction.

$$\eta = p_N m_o. \quad (18b)$$

#### Amount of Load Carried by One Tooth

According to Wellauer (ref. 5), the maximum load carried on one tooth of an uncrowned helical gear is determined from the load sharing ratio  $m_N$ . This is defined as the ratio of the face width  $F$  to the minimum total length of the lines of contact  $s_T$ :

$$m_N = \frac{F}{s_T} \quad (19)$$

$$\text{where } s_T = \sum_{N'=1}^{N_L} s_G = \text{minimum total length of lines of contact.} \quad (19a)$$

The above formula assumes that the load is constant along the length of the line of contact, which is a rather questionable assumption since the tooth stiffness is not constant at every point.

When analyzing crowned teeth, the author (ref. 6) has shown that the load sharing ratio can be expressed as follows:

$$m_N = \left( \frac{\eta_1}{\eta_2} \right)^3 \quad (20)$$

$$\begin{aligned} \text{where } \eta_2^3 = \eta_1^3 + \sum_{k_n=1}^{\alpha} \sqrt{[\eta_1^2 - 4k_n p_N (k_n p_N + 2f)]^3} \\ + \sum_{k_n=1}^{\beta} \sqrt{[\eta_1^2 - 4k_n p_N (k_n p_N - 2f)]^3} \end{aligned} \quad (20a)$$

$k_n$  = positive integer, which takes on successive value from 1 to  $\alpha$  or  $\beta$ , generating all real terms in the series. Imaginary terms should be ignored.

This expression is based on the assumption of a semi-ellipsoidal load distribution along each line of contact as shown in figure 16. It is also based on the assumption that the maximum intensity of pressure on adjacent teeth is a function of the mismatch of the tooth surfaces under no load; that is, that the intensity of loading at point P in figure 17 is the same as at point L because points P and L on the mating teeth have equal separation.

#### Position of Load on Tooth Surface

Since the intensity of pressure on the tooth surface varies with the length and position of the line of contact, it is generally necessary to analyze the contact conditions at a series of points on the tooth surface to arrive at the point where conditions are most severe. Generally the contact pressure is greatest at or very near the pitch line. For this reason it has been common practice to calculate contact pressures at the pitch line and to use these values in formulas for determining the pitting resistance of gear teeth. However, when the tooth numbers in the pinion are small and the pressure angle on the tooth is low, undercut may result; and the maximum contact pressure for spur, helical, and bevel gears will occur in the root of the pinion rather than at the pitch line. On spur



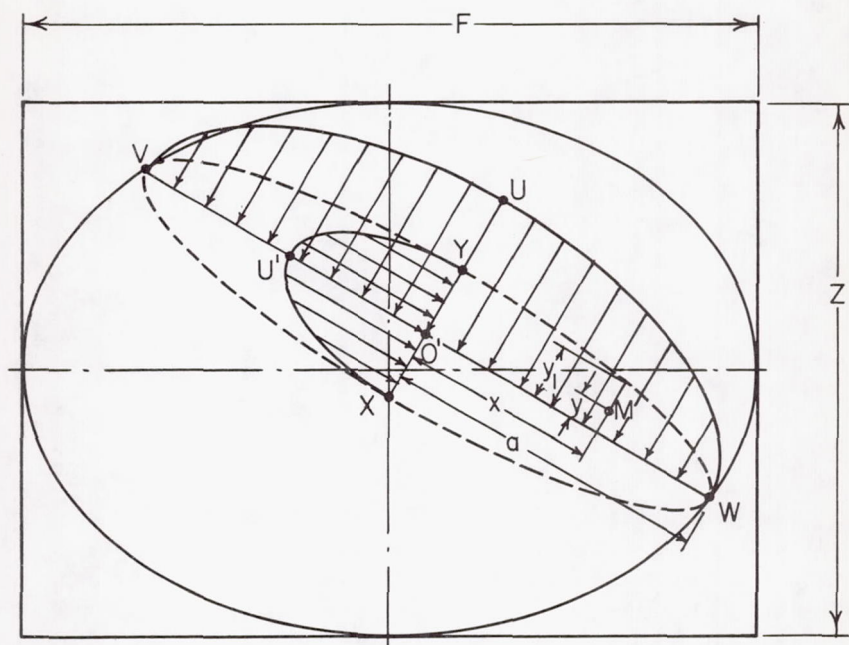


FIGURE 16.—Assumed load distribution along line of contact  $VW$ . This is an oversimplification.

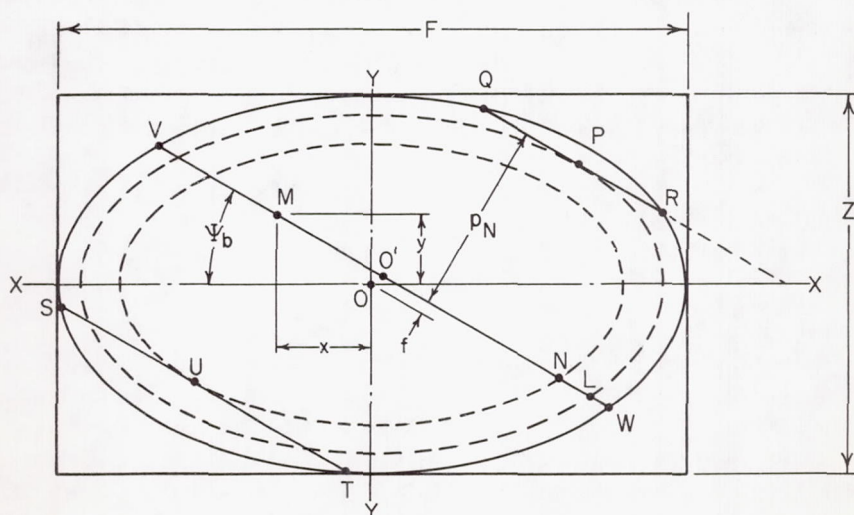


FIGURE 17.—The dotted ellipses are the secondary ellipses of constant separation.

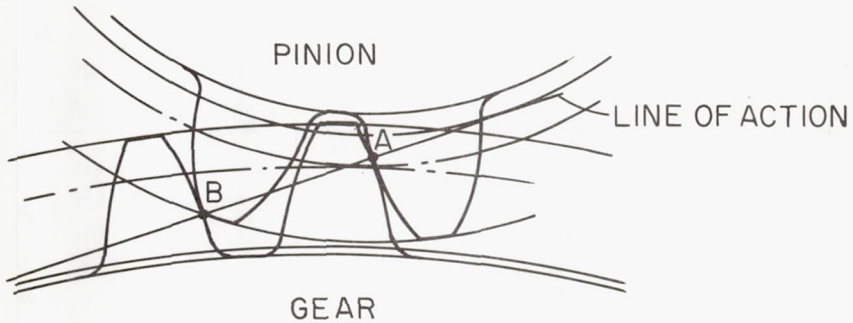


FIGURE 18.—The pinion tooth on the left is just coming into contact with the gear at point B. The point where the line of action crosses the left-hand profile of the right-hand pinion tooth (point A) denotes the lowest point of single tooth contact on the pinion.

gears and straight bevel gears, it may also occur at the lowest point of single tooth contact on the pinion as shown in figure 18.

There are other considerations besides maximum contact stress. Most formulas for wear and scoring depend upon both the pressure and the velocities at the point of contact. Since the pressure and velocity are seldom a maximum at the same point on the tooth surfaces, it is necessary to find the point on the tooth where the combination produces the most severe effect. For spur gears with no profile or lengthwise modification, it was long believed that scoring would be initiated at the tip of the pinion tooth where the sliding velocity was a maximum. More recently Kelley (ref. 7) showed that scoring could initiate at or near the highest point of single tooth contact on the pinion, since the pressure will be less at the tip of the tooth if load sharing occurs. Again it is necessary to analyze the contact conditions at a series of points to arrive at the point where conditions are most severe. Without a computer this can become a very tedious process.

For teeth that are crowned, it is almost essential to use an iterative technique to determine the most critical point. But before the critical point can be determined, it is necessary to have a failure theory.

#### Tooth Loads

Before discussing failure criteria, perhaps the most difficult aspect of gear design should be mentioned, i.e., tooth loads. The gear designer usually starts with a motor having a given rating. Frequently this rating does not represent the true capability of the motor. Electric motors are capable of brief overloads which can impart serious dynamic effects between the gear teeth. Then there is the nature of the driven machine which can impart additional overloads on the gear teeth. If these over-

loads occur rarely, they have little effect on pitting, scoring, and wear but may have a major effect on tooth breakage. If the overloads occur frequently, they will have a definite effect on pitting, scoring, and wear. Operating speeds coupled with gear tooth manufacturing inaccuracies can also have a pronounced effect on the loads transmitted to the gear teeth. Too many designers use arbitrary factors of safety, overload factors, and dynamic factors from standardized design practices. This frequently results in over-designed units or occasional under-designed units when conditions of operation have changed. It is surprising how little effort is made in most fields of design to obtain more reliable data concerning the nature of the tooth loads. Accurate appraisal of the loads is essential to efficient gear design.

#### Contact Stress

Once the dynamic tangential tooth load,  $W_d$ , has been determined, the equation for the contact stress,  $q_o$ , derived in reference 8 can be written:

$$q_o = C_p \sqrt{\frac{W_d L_m}{\rho_o s_G \cos \phi_n \cos \psi}} \quad (21)$$

$$\text{where } C_p = \sqrt{\pi \left[ \frac{k_c}{E_P} + \frac{1 - \nu_P^2}{E_P} + \frac{1 - \nu_G^2}{E_G} \right]} = \text{elastic coefficient.} \quad (21a)$$

$k_c$  = coefficient depending on contact conditions. For cylindrical contact (non-crowned teeth)  $k_c = 1$ . For spherical contact (crowned teeth)  $k_c = 1.5$ .

$\nu_P, \nu_G$  = Poisson's ratio for materials of pinion and gear, respectively.

$E_P, E_G$  = Young's modulus of elasticity for materials of pinion and gear, respectively.

$L_m$  = load distribution factor. This factor takes account of misalignment between the teeth due to errors in helix angle and shaft alignment in the gear mountings under load (ref. 8).

#### Traverse Distance Across Instantaneous Area of Contact

The distance through which the point of contact moves while crossing the area of contact depends upon the orientation of the area of contact and the direction of the path of contact (ref. 3).

To determine the time during which heat is generated at a given point on either member, refer to figure 10. On the pinion member, the direction of motion is represented by the direction of the absolute velocity vector,  $v_1$ . Since actual contact occurs over the area represented by the ellipse, heat will be generated during the time required for the point of contact



to move across the ellipse in the direction of motion. This distance is  $d_1$  in figure 10. In like manner, the absolute velocity vector for the gear is  $v_2$ , and the distance of travel is  $d_2$ . The inclination angle,  $w$ , in the tangent plane is determined as follows:

$$\tan w = \sin \phi_n \tan \psi \quad (22)$$

and the angles  $\alpha_1$  and  $\alpha_2$  which the resultant pinion and gear velocity vectors make with the pitch line are

$$\tan \alpha_1 = \frac{v_{pP}}{v_{FP}} \quad (23)$$

$$\tan \alpha_2 = \frac{v_{pG}}{v_{FG}} \quad (24)$$

The semiaxes  $a$  and  $b$  of the instantaneous ellipse of contact are

$$a = \frac{s_G}{2} \quad (25)$$

$$b = \frac{3}{\pi C_p^2} q_o \rho_o \quad (26)$$

Then,

$$d_1 = 2 \sqrt{\frac{a^2 b^2}{a^2 \sin^2(\alpha_1 + w) + b^2 \cos^2(\alpha_1 + w)}} \quad (27)$$

$$d_2 = 2 \sqrt{\frac{a^2 b^2}{a^2 \sin^2(\alpha_2 + w) + b^2 \cos^2(\alpha_2 + w)}} \quad (28)$$

For spur gears, the above formulas can be simplified to give

$$w = 0^\circ \quad (22a)$$

$$\alpha_1 = 90^\circ \quad (23a)$$

$$\alpha_2 = 90^\circ \quad (24a)$$

$$a = \frac{F}{2} \quad (25a)$$

$$b = \frac{3}{\pi C_p} \sqrt{\frac{W_d L_m \rho_o}{F \cos \phi_n}} \quad (26a)$$

$$d_1 = 2b \quad (27a)$$

$$d_2 = 2b \quad (28a)$$

#### Other Gear Types

The above formulas are sufficiently general to be used on spur, helical, straight bevel, spiral bevel, and hypoid gears. Other gear forms such as worm gears and crossed-helical gears require a similar treatment.

Worm gears are one commonly used form of crossed-axis gearing upon which relatively little theoretical analysis has been made. To be sure, the geometry is somewhat complex, but it is not beyond analysis. This is one field of gearing where further study is necessary and should be pursued. Present rating formulas are empirical and are entirely inadequate if one wishes to depart from the conventional speed reducer designs.

#### Surface Finish

Surface finish plays a very important role in the analysis of gear tooth wear and scoring. It influences the height of the asperities on the tooth surface, and it has a significant effect on the coefficient of friction. The original surface finish on a gear tooth may determine whether the gears can survive the break-in period without failure. It also frequently imposes a limit on the optimum surface finish that can be achieved during break-in.

It has recently been observed in tests conducted at IIT Research Institute on a privately sponsored program that surface finish has a significant effect on surface fatigue. Details of this work are not presently available.

#### Surface Micro Geometry

Another aspect of tooth surfaces is the micro geometry. This is the geometrical imperfection machined into the gear surfaces, such as generating flats and cutter flats, which are not normally considered when measuring surface finish. The intersection of two flats results in a ridge on the surface which can increase the actual surface stresses above the calculated surface stresses based on the macro geometry.

With a knowledge of the cutting machine and the cutter setup, it is theoretically possible to calculate the geometry of these flats. However, every cutting process produces a different set of formulas. As far as the author knows, no one is attempting to make these calculations. In fact, this micro geometry is generally ignored in gear design which may explain one of the causes for disagreements between various investigators.

#### Break-In

The break-in process is generally a haphazard one for most gear applications. Generally no particular thought is given to whether a break-in procedure is required. This is quite frequently due to the fact that the designer is not aware of the need for or use of a break-in; however, a break-in procedure can be quite useful. Its purpose is to condition the

gears for their future service. It can serve two useful functions: to improve the surface finish and to increase the contact area between mating teeth.

On unhardened gears, the surface finish can be improved by running the gears under load at a sufficiently low speed to effectively prevent wear and scoring as seen in figure 8. By gradually increasing the load, the asperities on the teeth will be plastically deformed, thereby reducing their height. In addition, the gross surface of the teeth may deform to produce mating surfaces which depart from the original shape. This may be either beneficial or disastrous. In the process some materials work-harden and thereby improve the durability of the surface.

In a similar manner the surfaces on hardened gear teeth may also be improved. However, because the surfaces are hard it is more difficult to deform them plastically. By operating at slightly higher speeds wear will occur, which should speed the break-in process. But this is a slow process at best. Still higher speeds will result in rapid wear and the likelihood of scoring.

The break-in procedure is essentially a process of operating under boundary lubrication without causing failure by excessive wear or scoring. The object is to smooth the surfaces sufficiently that hydrodynamic lubrication can exist under normal operating loads and speeds.

#### Surface Treatment

Certain surface treatments have been applied to hardened gears, such as the phosphate etching treatment. This process presumably makes the tooth surfaces porous, which allows said surfaces to absorb lubricant, thereby increasing the efficiency of boundary lubrication. The porosity of the surfaces may also make them more easily deformed plastically and therefore speed the break-in. Other methods of surface treatment include copper plating, silver plating, and tungsten disulphide coatings. These treatments have met with varying degrees of success. The only one in common use at present is the phosphate treatment. There is very little literature on the subject of surface coatings for gears. Further work in this field may be well justified. By proper choice of lubricants, surface treatments, and actual break-in procedure, it is possible to achieve a considerable increase in gear capacity.

#### Surface Hardness

It is known that an increase in surface hardness will decrease the likelihood of wear and scoring. It is not known for certain why this is true. The author believes that this may be due to the fact that, under boundary lubrication, the hard surfaces will operate at a higher temperature than softer surfaces. This is certainly an area that needs more study, since most of the available information concerns the wear of unlubricated surfaces



or lubricated surfaces in pure sliding and, therefore, is not applicable to the solution of gear problems where both rolling and sliding occur.

#### Coefficient of Friction

Enough experimental work has been performed to establish the fact that the coefficient of friction between gear teeth is not constant. Studies made by Kelley, Benedict, and Leach (refs. 9 and 10) on the geared roller test machine have shown that the coefficient of friction is both speed and load dependent, and that it varies through a range of at least one order of magnitude. These authors have shown that the coefficient of friction  $\mu$  can be expressed in the form

$$\mu = K_1 \log K_2 [v_s v_T^m \eta_o^n w^p] \quad (29)$$

where  $K_1$  and  $K_2$  are constants;  $m$ ,  $n$ , and  $p$  are exponents for a given lubricant;  $v_s$  is the sliding velocity;  $v_T$  is the sum of the two surface velocities  $v_1$  and  $v_2$ ;  $\eta_o$  is the viscosity of the bulk oil; and  $w$  is the unit load. Furthermore, the friction is somewhat dependent upon the surface finish. Figure 19 shows one typical example of the dependence of the coefficient of friction on load at a constant speed using mineral oil as determined by the author (ref. 11). This graph was obtained during a routine scoring test on rollers and is therefore typical of the results to be expected on gear teeth. No special break-in procedure was used.

There are still many unanswered questions concerning the variation in coefficient of friction. Certainly the coefficient can be expected to vary across the band of contact on crowned rollers. It is equally true that the coefficient can be expected to vary across the band of contact on gears.

Since the coefficient of friction at a point may well depend upon the micro geometry, it is not surprising that wear and scoring test results produce a large scatter. Furthermore, most of the friction experiments have been performed on a single roller geometry combination. It seems reasonable that the coefficient of friction will vary with the film thickness which in turn varies with the load, the velocities, and the contact geometry. This latter item has not received much consideration to date.

#### THE SOLUTIONS

Up to this point this paper has dealt with the definitions of various types of gear tooth surface failures and with the geometric and kinematic considerations in gear design which relate to these failures. The succeeding sections will touch briefly on the presently accepted or newly proposed methods for the solution to several of these types of failures.

#### Gear Tooth Pitting

The presently accepted method for designing against gear tooth pitting

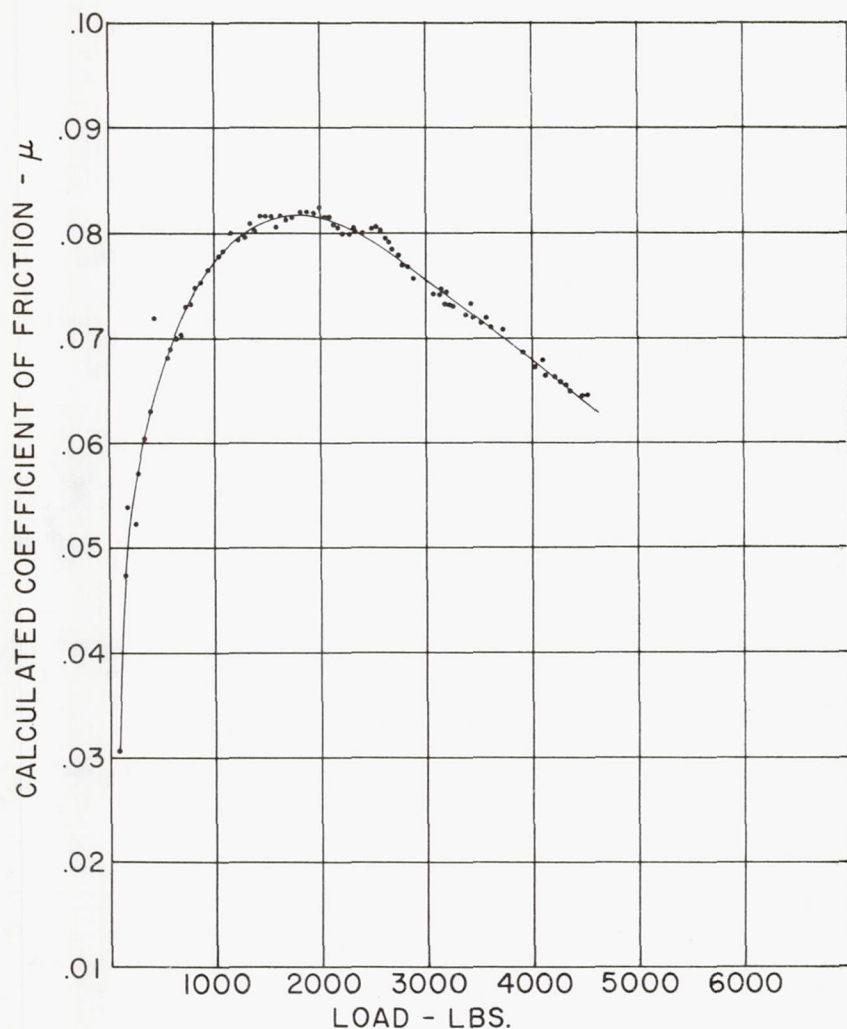


FIGURE 19.—Variation of the coefficient of friction with load between a 3.6-in. diameter roller (B) and a 2.4-in. diameter roller (A) on the geared roller test machine at 1000 rev/min of the smaller roller and a peripheral velocity ratio of 1.3125 between the larger and smaller rollers.  $v_A = 10.5$  ft/s,  $v_B = 13.8$  ft/s,  $v_S = 3.3$  ft/s.

is based largely on the solution to the Hertz equation for contact stress. Allowable limits of contact stress have been established for various gear materials and are in general use in gear design practices.

In 1935 Way (ref. 12) suggested that lubrication played a part in gear tooth pitting by showing how oil trapped in a small crack could exert very high pressures within the crack. His experiments with lubri-

cated and unlubricated rollers demonstrated that pits would propagate more rapidly in lubricated rollers.

Dawson (ref. 13) has shown that very small surface cracks may appear at a very early stage—20 percent of the life or less—and propagate rather slowly up to the point where a visible pit may appear. The work of Way and Dawson would appear to complement each other.

More recently another phenomenon has been observed concerning the effects of lubrication on pitting. While running contact fatigue tests on rollers at IIT Research Institute (ref. 14) using MIL-L-2105 lubricant, it was noted that, as the oil aged and the EP additives became depleted, the pitting resistance of the rollers increased. Tests performed by other investigators including Bailey and Beaubien (ref. 15) have confirmed the fact that the additives in EP lubricants usually reduce the pitting life appreciably. Certain EP lubricants will reduce the life by at least one order of magnitude over that of a straight mineral oil. Certainly there is a need for a further study of this phenomenon.

#### **Gear Tooth Wear**

Formulas for gear tooth wear seem to be non-existent. Although the author knows that some experimental work has been done on gear tooth wear, he is not aware of any papers specifically dealing with the subject. The gear engineer usually assumes that, if the gears have been designed to resist pitting, scoring, and breakage, there will be no problem from wear. This assumption is, no doubt, based on the fact that wear is generally a slow process of failure and that relatively infrequent inspections of the gear unit will uncover any possible trouble long before it has reached serious proportions. In the future this may not be a valid assumption. Some applications, such as those concerned with our space program, may not lend themselves to periodic inspections at sufficiently frequent intervals to correct the trouble of worn gears.

The author is familiar with applications in which heavily-loaded gears have been seriously affected by wear; therefore, formulas are needed. Perhaps the solution is to be found in the formulas for oil film thickness.

#### **Oil Film Thickness**

It is not the intent of this paper to duplicate the subjects covered by other authors at this symposium. Therefore, no attempt will be made to give formulas for oil film thickness. But the concept of an EHD oil film between gear teeth is relatively new to the gear engineer, and he has really made very little attempt to study its effects. Up until the 1960's, most designers believed that gears were lubricated by boundary lubrication in which the oil film thickness was less than or approximately equal to the asperity height. It is now believed by many that the oil film thickness may be much greater than the asperity height for many appli-



cations. Much work has been done in studying oil film thickness by Martin (ref. 16), Crook (ref. 17), Dowson and Higginson (ref. 18), and many others.

A few years ago it was quite commonly believed that gears would immediately start to wear out and that they would continue to wear until failure occurred. More recently engineers have become aware of gears on which the original cutter scratches are still visible after many years of service. This latter condition could only occur on gears lubricated with a full fluid oil film. This new awareness to the absence of wear is making the engineer more concerned with the theory of gear lubrication. Eventually it is probable that film thickness formulas will be accepted as a criterion for gear design. However, it is too early for these formulas to be accepted until the true relationship between film thickness and gear tooth wear can be confirmed through actual trials.

One point should be emphasized. Work to date has dealt primarily with the analysis of gearing with line contact such as noncrowned spur gear teeth. Very little work has been done to solve the problems of film thickness on gears with point contact (crowned teeth). Snidle and Archard (ref. 19) have made an attempt at a solution on hypoid gears with crowned teeth. Their theoretical results show an increase in film thickness with an increase in hypoid pinion offset, a result that is difficult to accept in view of the lubrication problems generally associated with hypoid gears. Since a high proportion of heavily-loaded gears are designed with crowned teeth, it is essential that the proposed formulas for gear film thickness take this into account.

#### **Gear Tooth Scoring**

Because scoring has been more of a problem in gear design than wear, especially in high speed applications, there have been many attempts to produce an adequate scoring resistance formula. Almen and Straub (refs. 20 and 21) proposed an empirical method for predicting scoring resistance of automotive helical and spiral bevel gears known as the PV-method. Later Almen (ref. 22) modified this earlier formula to take into account temperature effects. This latter method known as the PVT-method was applied principally to spur and helical gears, especially in the aircraft industry during World War II.

In 1937 Blok (refs. 23 and 24) made a theoretical study of the heat-flow equations at the point of contact between gear teeth. From his study he proposed a "flash temperature postulate"; i.e., for a given nonreactive mineral oil there would be a certain "critical temperature" above which the oil film would fail to give protection against scoring.

Gatcombe et al. (ref. 25) and Borsoff and Godet (ref. 26) have each developed a formula based on a time function of tooth contact. These formulas have not received widespread acceptance among gear engineers.

The film thickness formulas have also been proposed as a means for determining the failure point by scoring. With the latest EHD film thickness formulas, it is now believed that scoring will occur when the thickness of the oil film is less than the surface asperity height. Mention has been made previously of the work which has been done on film thickness theory.

#### Critical Temperature

One of the most promising formulas developed to date is the one which was proposed by Blok. Considerable work has been done by Kelley (ref. 7), Benedict and Kelley (ref. 9), Leach and Kelley (ref. 10), O'Donoghue and Cameron (ref. 27), and the author (refs. 3 and 11) using the Blok postulate. Presently Southwest Research Institute is conducting an industry sponsored program to evaluate more thoroughly the implications of the EHD lubrication theory and the critical temperature concept.

The critical temperature theory is being applied by gear engineers to the design of bevel and hypoid gears. It has been applied to the design of spur and helical aircraft gears as outlined by Kelley and Lemanski (ref. 28). There is fairly strong evidence that this theory does work with nonreactive oils.

In general, where the calculated film thickness is adequate to provide hydrodynamic lubrication, the critical temperature will be sufficiently low to be within safe limits. For this reason there has been considerable discussion as to which of these two criteria is the correct one for evaluating scoring resistance. When a hydrodynamic oil film exists, the coefficient of friction will be relatively low, and therefore the frictional heating in the contact area will be low. When boundary lubrication occurs, the asperity contacts will increase the coefficient of friction, and the frictional heating will increase.

However, the author (ref. 11) using the critical temperature theory and Snidle and Archard (ref. 19) using the film thickness theory find that, in the analysis of hypoid gears with crowned teeth, the two theories produce different results. With an increase in pinion hypoid offset, the calculated conjunction temperature increases, whereas the calculated film thickness also increases. This seems to be a contradiction. More work is required to determine the correctness of the two theories.

Too little work has been performed with reactive type lubricants to determine whether either of these two theories will properly evaluate scoring resistance. However, in the analysis of reactive type lubricants it is possible that the EP additives may have an effect similar to that of a preliminary surface treatment but in a milder form.

Two major problems confront the investigators of this theory—the difficulty of measuring the gear blank temperature near the point of contact and the unsolved problem of determining the instantaneous coefficient of friction at the point of failure just prior to scoring.



### TESTING TECHNIQUES

In order to evaluate these theories, it is necessary to perform tests under controlled operating conditions. The most common type of test is that using actual gears. Another frequently used method employs a disk machine in which two circular rollers are geared together to produce controlled sliding. Both of these tests are useful.

#### Gear Test Machine

The gear test machine has the advantage that it employs actual gears in a typical gear-box environment. Various lubricants can be operated in the same environment and can therefore be compared with one another under actual gear operating conditions. The three most popular gear machines for lubricant testing are the Ryder gear tester (U.S.A.), the IAE gear rig (U.K.), and the FZG gear rig (Germany). Descriptions and comparisons between these machines are quite thoroughly covered in reference 29.

Surprisingly these three machines do not rate lubricants in the same order. The gears used in each differ in size and tooth numbers, and operating characteristics also differ.

The principal disadvantage of the gear test rig is the fact that the contact pressures and velocities differ at every point through the meshing cycle. It is therefore difficult to determine under what instantaneous conditions the failure occurs, and there is no way to determine experimentally the order of the variables in the proposed formula. On the other hand, the gear test machine is the ideal means for determining whether a given formula does indeed evaluate the scoring resistance of a pair of gears.

#### Roller Test Machine

The roller test machine has the advantages that (1) the roller geometry is simple and easily controlled, (2) the contact pressures and velocities are known and under the control of the test engineer, and (3) the cost of the parts is much less than the cost of gears. It provides the ideal medium for experimentally determining the order of the variables. Furthermore, on the Caterpillar geared roller test machine it is possible to measure the friction torque during each test, thereby making it possible to determine the average coefficient of friction across the contact band. This machine is used to simulate the contact conditions occurring between gear teeth, though it cannot completely replace tests on actual gears.

#### Other Test Methods

The oil industry uses many other methods for evaluating lubricants, but the ones most commonly used for evaluating the wear properties of



lubricants are the pin on disk machines, the four-ball tester, and the Falex tester. Since these tests deal primarily with sliding contacts, they are not too suitable for evaluating gear lubricants where sliding is combined with rolling. This combination of sliding and rolling creates a different lubricant environment that needs to be considered.

The gear test machine and roller test machine do combine both rolling and sliding. These methods result in establishing more suitable gear lubricant ratings.

#### LUBRICANTS

Little has been said so far about gear lubricants. These can be grouped in three classifications: mineral oils, synthetic oils, and EP lubricants.

Mineral oils are widely used in industrial applications where the operating environment leads to reasonably constant temperatures. Since most antifriction bearings are best lubricated by a straight mineral oil, this means that a common lubrication system for gears and bearings is natural.

Synthetic lubricants have been developed to operate at temperatures beyond the normal range for mineral oils. The latter are too viscous at low temperatures and decompose at high temperatures. For military purposes a greater operating temperature range was required to achieve successful operation of equipment in both the arctic and the tropics with one lubricant. Synthetic lubricants generally possess a relatively low viscosity and will not provide thick film lubrication under as wide a range of loads and speeds.

Extreme pressure lubricants are a combination of EP additives in a mineral or synthetic oil. These additives have a corrosive effect on most metallic surfaces and therefore cause more rapid wear. But they do reduce the tendency toward welding or scoring by providing a surface coating that contaminates the two contacting surfaces. EP lubricants are undesirable for the lubrication of anti-friction bearings.

Many EP additives do not become active at low temperatures and tend to decompose more rapidly at high temperatures. They, therefore, generally have a rather narrow operating temperature range.

From the gear designer's point of view, the ideal gear lubricant is one that does not increase wear, will operate over a very wide temperature range, will have good viscosity characteristics over the whole temperature range, will provide anti-scoring characteristics, and will have good chemical stability, freedom from sludge formation, and be insensitive to moisture. Such a lubricant should be capable of lubricating both gears and bearings.

#### CONCLUSIONS

Much fine theoretical work has been done in an attempt to solve the problems of gear surface failure, but much remains to be done. The author has attempted to point out some of the areas where further investi-

gation is required and has listed the pertinent formulas to define the geometry and kinematics of gear design so that the investigator will have a better understanding of the actual conditions occurring in gear tooth contact. This is one side of the problem that is often overlooked when the theory is being developed and the test procedures are being established.

In conclusion it would be well to summarize the key problems facing the gear designer. These are some of the areas that need further study:

(1) A better method for accurately assessing the loads on the gears. It is not uncommon for the actual loads to differ from the design loads by a factor of two.

(2) Better design formulas for crossed-axis gearing such as worm gears and crossed-helical gears. Very little analytical study has been made of these types.

(3) Better methods for measuring surface finish on gear teeth are required. Present methods of measurement are inadequate.

(4) Micro geometry needs greater attention. This characteristic has been ignored up to now.

(5) The effects of various break-in procedures have received little attention. There may be positive advantages with a proper break-in routine.

(6) Surface treatments have been used successfully on gears. Our knowledge of these beneficial mechanisms leaves room for further study.

(7) The effects of surface hardness on lubricated gears are generally recognized. However, a full evaluation has not been undertaken, and the true reasons are not known.

(8) The measurement of the coefficient of friction in gear teeth has never been achieved. Efforts to measure the coefficient of friction between rollers has resulted only in an average value across the band of contact. There is much more work to be done here.

(9) It has long been known that pitting failure propagates more rapidly on lubricated surfaces. It now appears that variations in lubricant properties have a major effect on pitting failure, but little work has been done so far to determine the relationship.

(10) Recently it has been found that the surface finish on rollers with combined rolling and sliding contact has a pronounced effect on pitting resistance. Confirmation and further evaluation is necessary.

(11) Formulas for contact fatigue and scoring resistance of gears have been proposed. A formula for wear has not received much attention. To complete the analysis this omission must be rectified.

(12) There are presently two theories to explain gear tooth scoring: oil film thickness and critical temperature. For the gear designer a resolution of the overall problem is required.



(13) Much of the testing to prove out theoretical models is performed on rollers and other simple elements. Correlation with actual gear tests is required.

The above list should assure people that not all of the problems have been solved.

#### DISCUSSIONS

**P. M. Ku (Southwest Research Institute, San Antonio, Texas)**

Mr. Coleman has presented a comprehensive and concise review of the design considerations for spur, helical, and hypoid gears that have relatively similar contact conditions. He has also emphasized the need for a more analytical approach to the design of gears with very different contact conditions, such as worm and cross-helical gears. There is little doubt that this glaring gap in gear design urgently needs to be bridged.

He has discussed the implications of the elastohydrodynamic lubrication theory on wear and the critical temperature hypothesis on scuffing, suggesting that "a resolution of the overall problem is required." My feeling is that the two concepts are not incompatible. As long as a full EHD film prevails, there can be no asperity contacts and hence no wear or scuffing. In order for wear or scuffing to occur, asperity rubbing is a requirement but not necessarily the only requirement. The evidence is that whether the damage will take the form of wear or scuffing depends upon the balance between the frictional heat generation in the conjunction and the heat loss from the conjunction. At low sliding velocities, this thermal balance does not favor scuffing, thus only wear occurs. At high sliding velocities, the heat loss from the conjunction becomes less favorable, but the attainment of a critical thermal condition for scuffing depends also on the frictional heat generation in the conjunction. With straight mineral oils, not much asperity rubbing can be tolerated before the onset of scuffing, due to the high coefficient of friction. On the other hand, with reactive oils, the reduction in the coefficient of friction permits considerable asperity rubbing before scuffing occurs. In other words, the EHD theory provides a valid criterion for wear and a safe, though overly conservative, criterion for scuffing. The value of the critical temperature hypothesis is that it provides an extra margin for design against scuffing. This extra margin is generally small for straight mineral oils, but can be very substantial for reactive oils. The difficulty in applying the critical temperature hypothesis to reactive oils is that their critical temperatures are not constant. However, until a better thermal criterion is found, the critical temperature hypothesis is still a tangible guide which, if employed with caution, is most helpful in design.

Mr. Coleman's discussion on the effect of pinion offset for hypoid gears is most intriguing. Citing his own and Archard's results, he states that an increase in pinion offset increases both the calculated conjunction



temperature and the calculated film thickness. Two questions come to mind: First, if the conjunction temperature increases with increased pinion offset, does it mean that the actual load-carrying capacity decreases with increasing pinion offset? Second, the film thickness is dependent upon the sum velocity, not the sliding velocity. Does it mean that the vectorial sum velocity at the moving conjunction also increases with increasing pinion offset? Mr. Coleman's thoughts with respect to these two questions will be appreciated.

**E. E. Bisson (NASA Lewis Research Center, Cleveland, Ohio)**

I would like to ask Messrs. Kelley and Coleman for an objective evaluation of relative advantages and disadvantages of gear research using full-scale gears versus roller or disk machines. Some indication of the correlation between gear rigs and practical experience versus correlation between roller or disk rigs and practical experience would be information of considerable interest.

**B. W. Kelley (Caterpillar Tractor Company, Peoria, Illinois)**

For over 20 years the writer has conducted many tests on full scale transmissions and four-square test machines using both small and large gears. The tremendous costs of such tests were incurred because this is perhaps the only way in which to learn about kinematics, the response to failures, the pressure distribution, the need for geometric control, and the statistics of performance of such parts. But one is naive to feel that much can be learned about the problems of concentrated contact as affected by materials and lubricants by using such test equipment. The gear is a complex piece of machinery. Very modest and seemingly insignificant errors play a large role in the pressure distribution and the performance of such parts. For those people who need this kind of knowledge and are willing to devote many years to the analytical support of tests, full-scale gear test machines are meaningful.

On the other hand, if one wants to learn something about the problems of concentrated contact, he is well advised to avoid the complexity of performing tests on such complicated and costly pieces of equipment. The writer has always advocated the use of roller or disc machines for doing this work. With such equipment, one can easily duplicate almost any instantaneous kinematic situation that occurs on a gear. For those designers who are skilled in the analysis of the full-scale gear, such information coming from roller test equipment is extremely valuable. The correlation between roller test rigs and full-scale gear tests are good only when the analysis performed on the gears is sufficiently sophisticated to define the actual pressures and bulk temperatures of the operating parts. We have published some of this correlation (refs. 30 and 31). It is the writer's impression that many companies are beginning to develop a

much higher analytical capability in their gear trains. More and more recognition of the value of performing tests on the simpler equipment will result and less money will be spent on costly full-scale equipment where conditions cannot be so well defined.

#### LECTURER'S CLOSURE

Mr. Bisson has asked for an objective evaluation of the relative advantages and disadvantages of a roller test machine vs a full-scale gear test rig.

The gear test rig has the advantage of testing the product in its own environment. The overall effects of gear housing deflections on tooth misalignment, of the method of applying the lubricant, and of the varying surface pressures and sliding velocities will be incorporated in the results. However, it is nearly impossible to determine the specific effects of each. This has therefore led to the use of the roller test rig.

The roller test rig provides the following advantages:

- (1) Constant sliding and rolling velocities of the two rollers
- (2) Constant tooth contact pressure, which can readily be controlled by roller crowning and by variation in the applied load
- (3) Constant surface geometry of the test rollers
- (4) Better control of surface finish on the test surfaces
- (5) Constant conditions in the application of the lubricant, since the lubricant stream is uninterrupted by tooth slots.
- (6) Simpler and more accurate measurement of surface temperature on the specimens
- (7) Greater ease in measuring bearing losses on the machine
- (8) Constant known operating conditions, such that each parameter can be varied and evaluated independently
- (9) Simpler and therefore less expensive test specimens

It is this ability of the roller test machine to evaluate the independent effects of each variable that makes it a most useful tool. The gear test rig does not have this capability. When correlating the results between laboratory tests (either gear test rig or roller test rig) and actual gear applications, the following comments may give some insight.

On bevel gears the critical temperature hypothesis has been in use for approximately 10 years to evaluate the susceptibility to scoring. To date it has been reasonably successful in predicting the likelihood of scoring on critical bevel gear drives. It has been used with a moderate degree of success with reactive type lubricants used on hypoid gears. However, the greatest problem in determining correlation between roller tests and practical gear applications is the lack of accurate data concerning actual tooth contact geometry including profile, lead, and spacing errors; gear blank temperatures; tooth loads and speeds; and dynamic responses from



the overall gear drive. Where all of these factors are known, correlation has been reasonably good.

By using roller tests, data can be obtained relatively inexpensively. Ultimately a limited number of gear tests are performed to validate the data. These gear tests are performed under controlled conditions.

The author agrees with Mr. Ku concerning the fact that the film thickness and critical temperature concepts may be compatible. But we really do not know all of the answers, and further study is required.

Concerning the effects of hypoid offset on scoring, the calculated conjunction temperature increases with an increase in the offset. This implies a decrease in load-carrying capacity with an increase in pinion offset, which has been borne out by actual hypoid gear experience. This conclusion appears to be contrary to the finding of Snidle and Archard (ref. 19), which is based on calculations of minimum EHD film thickness. These latter calculations show that the film thickness increases with an increase in offset. Therefore, it would appear that the load-carrying capacity should increase with an increase in the pinion offset.

Calculations of the sum velocity on hypoid gear teeth show an increase of 56 percent for an 8.875-inch diameter hypoid gear with a 2.500-inch pinion offset over that of the corresponding spiral bevel gear. This would seem to imply an increase in film thickness. However, the sliding velocity increases by 410 percent. It is the effect of the sliding velocity on film thickness that seems to be in doubt.

#### NOMENCLATURE

$A$	mean cone distance of bevel or hypoid gear, in.
$a$	semiaxes of instantaneous ellipse of contact, in.
$b$	
$C_p$	elastic coefficient, $\text{lb}^{1/2}/\text{in.}$
$d$	length of contact in direction of sliding, in.; [ $d_1$ and $d_2$ refer to values for pinion and gear, respectively]
$E$	Young's modulus of elasticity, $\text{lb}/\text{in}^2$ ; [ $E_P$ and $E_G$ refer to values for pinion and gear, respectively]
$F$	face width, in.
$f$	distance from center of surface of action to line of contact, in.
$K$	distance along path (line) of action from start of engagement to line of contact, in.
$K_1$	constants used in equation for coefficient of friction
$K_2$	
$k_c$	coefficient depending on contact conditions
$k_n$	positive integer
$m$	exponent in equation for coefficient of friction



$m_F$	face contact ratio
$m_N$	load sharing ratio
$m_o$	modified contact ratio
$m_p$	transverse (profile) contact ratio
$N'$	positive integer
$N_L$	minimum number of lines of contact in a gear mesh
$n$	exponents in equation for coefficient of friction
$p$	
$p_b$	
$p_N$	mean normal base pitch, in.
$q_o$	maximum contact pressure (Hertz contact stress), lb/in <sup>2</sup>
$R$	gear pitch radius, in.
$r$	pinion pitch radius, in.
$s_G$	length of the line of contact on one tooth, in.
$s_T$	minimum total length of the lines of contact in the gear mesh, in.
$v$	velocity of point of contact, in/sec; [ $v_1$ and $v_2$ refer to the values of resultant velocity vectors on pinion and gear, respectively]
$v_F$	lengthwise sliding velocity, in/sec; [ $v_{FP}$ and $v_{FG}$ refer to the values of lengthwise sliding velocities on pinion and gear, respectively]
$v_n$	tangential velocity at pitch point in normal plane, in/sec
$v_n'$	tangential velocity at point $Q$ in normal plane, in/sec
$v_p$	profile sliding velocity, in/sec; [ $v_{pP}$ and $v_{pG}$ refer to the values of profile sliding velocities on pinion and gear, respectively]
$v_s$	resultant sliding velocity, in/sec
$v_T$	sum of the two surface velocities $v_1$ and $v_2$ , in/sec
$W_d$	dynamic tangential tooth load, lb
$w$	unit tooth load, lb/in.
$w$	inclination angle between line of contact and the pitch line in the tangent plane
$Z$	length of action in transverse section, in.
$z_o$	distance along line of action in mean normal section from the pitch point to the critical point for pitting, scoring, or wear, in.
$\alpha$	angle which resultant velocity vector makes with pitch line; [ $\alpha_1$ and $\alpha_2$ refer to values for pinion and gear, respectively]
$\gamma, \Gamma$	pitch angles of pinion and gear, respectively
$\Delta$	crowning between teeth, in. This is the total separation between the pinion and gear tooth surfaces at the ends of the teeth.
$\eta_1, \eta_2$	load sharing coefficients, in.
$\eta_o$	viscosity of bulk lubricant
$\mu$	coefficient of friction
$\nu$	Poisson's ratio; [ $\nu_P$ and $\nu_G$ refer to values for pinion and gear materials, respectively]

$\rho$	radius of profile curvature, in.; [ $\rho_1$ and $\rho_2$ refer to values of the radii of profile curvature at the point of contact in the mean normal section on pinion and gear, respectively. $\rho_P$ and $\rho_G$ refer to values of the radii of profile curvature at the pitch point in the mean normal section on pinion and gear, respectively].
$\rho_L$	relative radius of lengthwise curvature, in.
$\rho_o$	relative radius of profile curvature, in.
$\phi_n$	normal pressure angle
$\psi$	mean helix (spiral) angle
$\psi_b$	base helix (spiral) angle
$\omega_P$	pinion angular velocity, rad/sec

## REFERENCES

1. LEWIS, W.: Investigation of the Strength of Gear Teeth. Engineers' Club, Philadelphia, Pa., October 15, 1892.
2. DUDLEY, D. W.: Elastohydrodynamic Behavior Observed in Gear Tooth Action. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965-66, p. 157.
3. COLEMAN, W.: A Scoring Formula for Bevel and Hypoid Gear Teeth. Trans. ASME, J. Lub. Tech., vol. 89F, 1967, p. 114.
4. BAXTER, M. L.: Basic Geometry and Tooth Contact of Hypoid Gears. Industrial Mathematics, vol. 11, pt. 2, 1961, p. 19.
5. WELLAUER, E. J.: Helical and Herringbone Gear Tooth Durability—Derivation of Capacity and Rating Formulas. AGMA 229.06, June 1962.
6. COLEMAN, W.: Improved Method for Estimating the Fatigue Life of Bevel and Hypoid Gears. SAE Quarterly Trans., vol. 6, 1952, p. 314.
7. KELLEY, B. W.: A New Look at the Scoring Phenomena of Gears. SAE Trans., vol. 61, 1953, p. 175.
8. COLEMAN, W.: Pitting Resistance of Bevel and Hypoid Gear Teeth. AGMA 229.05, Oct. 1960.
9. BENEDICT, G. H.; AND KELLEY, B. W.: Instantaneous Coefficients of Gear Tooth Friction. ASLE Trans., vol. 4, 1961, p. 59.
10. LEACH, E. F.; AND KELLEY, B. W.: Temperature—The Key to Lubricant Capacity. ASLE Trans., vol. 8, 1965, p. 271.
11. COLEMAN, W.: Bevel and Hypoid Gear Surface Durability: Pitting and Scuffing. Proc. Inst. Mech. Engrs., vol. 182, pt. 3A, 1967, p. 191.
12. WAY, S.: Pitting Due to Rolling Contact. J. Appl. Mech., vol. 2, 1935, p. 49.
13. DAWSON, P. H.: Rolling Contact Fatigue Crack Initiation in a 0.3 Per Cent Carbon Steel. Proc. Inst. Mech. Engrs., vol. 183, 1968.
14. ANON.: The Surface Fatigue of Gear Steels. IITRI Rept. B8012-7, Aug. 1965, p. 23.
15. BAILEY, D. R.; AND BEAUBIEN, S. J.: Study of Helicopter Gear Lubrication. Contract N00019-67-C-0549, Naval Air Systems Command, Dept. of Navy, 1968.
16. MARTIN, H. M.: Lubrication of Gear Teeth. Engineering (London), vol. 102, p. 199.
17. CROOK, A. W.: Simulated Gear-Tooth Contacts: Some Experiments Upon Their Lubrication and Subsurface Deformations. Proc. Inst. Mech. Engrs., vol. 171, 1957, p. 187.

18. DOWSON, D.; AND HIGGINSON, G. R.: Elastodynamic Lubrication. Pergamon Press, 1966.
19. SNIDLE, R. W.; AND ARCHARD, J. F.: Lubrication of Elliptical Contacts. Inst. Mech. Engrs. Tribology Convention, Gothenburg, Sweden, May 1969.
20. ALMEN, J. O.: Durability of Spiral-Bevel Gears for Automobiles. Automotive Ind., Nov. 16 and 23, 1935.
21. ALMEN, J. O.; AND STRAUB, J. C.: Aircraft Gearing: Analysis of Test and Service Data. AGMA, 1948.
22. ALMEN, J. O.: Dimensional Value of Lubricants in Gear Design. SAE J., vol. 50, 1942, p. 373.
23. BLOK, H.: Les temperatures des surface dan des conditions de graissage sous pression extreme. Congr. mondial petrole, 2 me Congr., Paris, vol. 3, 1937.
24. BLOK, H.: Measurement of Temperature Flashes on Gear Teeth under Extreme Pressure Conditions. Proc. Gen. Disc. on Lubrication, Inst. Mech. Engrs., vol. 2, 1937, p. 14.
25. GATCOMBE, E. K.; HUNNICUTT, R. P.; AND KINNEY, G. S.: Spur Gears—Hertzian Contact Times. Lub. Engrg., vol. 16, 1960, p. 308.
26. BORSOFF, V. N.; AND GODET, M. R.: A Scoring Factor for Gears. ASLE Trans., vol. 6, 1963, p. 147.
27. O'DONOGHUE, J. P.; AND CAMERON, A.: Temperature at Scuffing. Proc. Inst. Mech. Engrs., vol. 180, pt. 3B, 1965-66, p. 48.
28. KELLEY, B. W.; AND LEMANSKI, A. J.: Lubrication of Involute Gearing. Proc. Inst. Mech. Engrs., vol. 182, pt. 3A, 1967, p. 173.
29. ANON.: Gear Lubrication. Inst. of Petroleum, Elsevier Publ., Amsterdam, 1966.
30. DENNING, R. E.; AND RICE, S. L.: Surface Fatigue Research with the Geared Roller Test Machine. SAE Trans., vol. 72, 1964.
31. DENNING, R. E.: Gear Material, Heat Treatment, and Lubrication Research with the Geared Roller Test Machine. AGMA Meeting, Oct. 1964.